Universitá degli Studi dell'Insubria
Dottorato di Ricerca in Matematica del Calcolo:
Modelli, Strutture, Algoritmi e Applicazioni
XXIII Ciclo

Extreme Value Stochastic Processes: Vasicek Model Revised

A Dissertation Submitted In Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

by

Stefano Hajek

Supervisor: Stefano Serra Capizzano
Full Professor, Universitá degli Studi dell'Insubria

Assistant Supervisor: Marco Donatelli
Assistant Professor, Universitá degli Studi dell'Insubria

A.A 2012/2013

Aknowledgements

A PhD experience can be a starting point but is undoubtedly an arrival; that's true in particular for me because before this experience I spent a lot of time (two decades) in practicing, and (partially) understanding the solutions that my job daily requires to find. For this reason it's quite impossible to mention all the contributions, suggestions, ideas and all the people that helped me, directly or indirectly, in obtaining this result. Hence, colleagues and friends that I'm going to remember here are representative of a larger set of interactions that make - under my only responsibility - this work as is.
Prof. Stefano Serra Capizzano, and dott. Marco Donatelli are the nearest in time that paid attention to this effort and the first I'm grateful to.
A special regard goes to prof. Michele Donato Cifarelli that gave me an example of scientific devotion out of mere academic duty.
I'm honoured to cite prof. Domenico Candeloro and prof. Franco Cazzaniga - that followed me in my preparation - as good friends of mine so as prof. Giulianella Coletti that first encouraged me in my research intention.
Some regret comes from my impossibility to live the every day life at the Math Dept. and to spend more time with the other PhD students; fortunately this fact didn't prevent a pleasant relationship with most of them.
This course has been - despite isolated episodes - a definitive evidence of the importance of a genuine spirit of cooperation in achieving a personal goal while pursuing a commom interest; that's the fundamental lesson I'd like to testify with this work.

Contents

Chapter 1

Overview: extreme values in "shocked" economies

Risk management is nowadays an absolute priority for financial institutions not only as a consequence of actual market crisis and default warnings but as a systematic prevention against undesired shortfall of complex speculative instruments.
Basel 3 as well Solvency 2 international regulatory framework is the main evidence of a generalized interest in loss mitigation and capital protection.

From a technical point of view risk management implies a trade between certain and uncertain quantities and makes use of advanced asset "insurance" strategies; insurances have a cost depending from the payoff expected by the insurer to take risks.

In order to assess correctly the costs of risk exposures risk managers make use of quantitative models based on standard assumptions like log normal distribution of assets prices; current stressed economy definitively confirms a well known drawback: this class of models cannot explicitly account for the negative skewness and the excess kurtosis of asset returns.
Empirical studies show that underlying's prices distribution is NOT lognormal; as noted by Jackwerth and Rubinstein [48] ,for example, in a lognormal model of assets prices, the probability of a stock market crash with some 28% loss of equity values is 10^{-160}, an event which is unlikely to happen even in the life time of the universe.
Current working experience in financial institutions definitively confirms that classical assumptions about the normality of return distribution are inadequate to represent financial scenarios. For example, functional parameters of interest rates models calibrated on recent historical series, take values out of their

validity range; again, forcing observations to fit gaussian distribution in Value at Risk calculations lead to dangerous prociclical effects: to take into account exceptional occurred events in a normal density estimation we have to stretch the volatility parameter so that, in the future, the risk will be overestimated and the related actions will be too cautious and ineffective. Maybe the misleading financial models are not the cause of economic recession but it is a matter of fact that wrong beliefs carry distortions in expectations and can compromise strategic decisions; awareness should be the first remedy for the crisis and misunderstandings aren't useful to develop awareness.

My attempt to build an interest rates model well suited for extreme returns is documented in this PhD dissertation; this attempt is based on well-established results in Extreme Value Theory, and applies them to stochastic processes changing the distributional assumptions of the Vasicek model. An experimental test verifies the effectiveness of this approach. Before any other enquiry this introductory chapter is dedicated to give some evidence about the relevance of the topic; fortunately the literature on empirical analysis about the violation of normality assumptions in financial markets is exhaustive and I can refer to it without replicate it. From a theoretical perspective it is interesting to understand the reasons why the financial series show a non-normal behaviour; so the main part of the chapter gives some insight about those reasons and lays basic concepts preliminary to the more formal exposition of the next chapters.

1.1 Microstructure models of market returns

Empirical analysis of stock returns exhibit their fat tailed distribution showing that extreme price movements are frequent while Gaussian distribution under evaluate them.

The explanation of non-Gaussian returns is a research field that includes famous contributions from Mandelbrot and Fama and, starting from them, evolves two main directions in understanding the non-Gaussian shape of the return distribution.

The first direction deal with the mixture-of distributions hypothesis, which states that return distributions are a mixture of Gaussian distributions with different variances [13]; under this hypothesis variance changes are induced by fluctuations in the rate of trade. A second view for the non-Gaussian shape of the return distribution is the stable Paretian Hypothesis: it argues that returns are drawn independently and identically from a stable or truncated stable distribution [29], [65].

In what follows an introduction to the stable Paretian and mixture of distribution hypotheses is given with some comment about their compatibility with the extreme value approach to interest rates dynamics that I'm going to propose.

Suppose to observe the stream of stock prices p_t $t = 1, ..., T$ during a trading day; the intraday returns are given by: $r_t = \log\left(\frac{p_t}{p_{t-1}}\right)$.

Under the central limit theorem we are legitimated to expect that the daily overall return $\varpi = \sum_{t=1}^{T} r_t$ is a random variable that follows a normal distribution.

This intuition is at the basis of the most celebrated quantitative financial models from the early paper of Louis Bachelier [1] that in 1900 introduced the random walk model for securities and commodities market in which the innovation terms of the stochastic process (see further for details) are independent and normally distributed.

1.1.1 The Stable Paretian hypothesis

In 1963 Benoit Mandelbrot observed that stock returns samples typically contain so many "outliers" that normal density fitted to the mean square of price changes are lower and flatter than the distribution of the data themselves; it means that the tails of the distribution of price changes are so long that the sample second moments vary in an erratic fashion.

Starting from this fact Mandelbrot suggests a new model of price behaviour that replaces the Gaussian distribution with the family of "stable Paretian" family of probability laws which were first described by Paul Levy in 1925 [61]; being the Gaussian a limiting case of the Levy family, the Mandelbrot's model is a generalization of Bachelier's one.

Levy defined and described stable distributions through the characteristic function.

If X_1 and X_2 are independent random variables drawn from a given *stable* distribution, then for some real constants c_1 $c_1$there must another real constant c such that $c_1X_1 + c_2X_2 = cX$ where X has the same distribution as X_1 and X_2

$$\text{Let } \varphi(u) = \begin{cases} \exp\left(i\delta u + \gamma^\alpha \left\{-|u|^\alpha + i\beta u |u|^{\alpha-1} \tan\left(\frac{\alpha\pi}{2}\right)\right\}\right) \\ \exp\left(i\delta u + \gamma^\alpha \left\{-|u|^\alpha + i\beta u \frac{2}{\pi} \log(|u|)\right\}\right) \end{cases}$$

denote the characteristic function, it follows that $\varphi(c_1\mu)\varphi(c_2\mu) = \varphi(c\mu)$ and that $c^\alpha = c_1^\alpha + c_1^\alpha$
(because this property, stable distributions are also referred as "alpha stable").
Setting $\alpha = 2$ and $\beta = 0$ in $\varphi(u)$ we find the characteristic function of the normal distribution that is a special member of the stable family, otherwise, setting $\alpha = 1$ and $\beta = 0$, $\varphi(u)$ becomes the characteristic function of the Cauchy distribution; inverting it analytically we have the density $f(x) = \frac{\gamma}{\pi(\lambda^2+x^2)}$ that has fatter tails than the Gaussian.
Hence, the variation of αparameter varies the tail of the distribution, while the β parameter, being associated with the relative importance of the two sides of the distribution, it determines the skewness of the distribution.
Mandelbrot's proposal, altough its relevance, didn't obtain a full aknowledgement because the practical shortcoming it is exposed: Levy distributions, not allowing for the existence of the variance, conflicts with classical quantitative finance and its standard solutions (e.g. the mean-variance portfolio models).
Furthermore, being unavailable a closed form solution for the density – the characteristic function is in general not analitically invertible – parameter estimation becomes a critical task, partially overcome by modern computing power that allows numerical inversion.

1.1.2 The Mixture of Distributions hypothesis

As previously mentioned, an alternative to the stable Parethian hypothesis is the "*mixture of distributions*" hypothesis originally proposed by Clark.
Clark analyzes the arrival rate of events suggesting that information speed is not constant, there are days on which more news is released than on others and that are consequenty expected to display greater volatility.
Consider the stochastic process $T(t)$ such that $t < s \Rightarrow T(t) \leq T(s)$ with $t \in \Re^+$and the process for log-price $v_t = \log(p_t)$

$$v_t = \mu T(t) + \sigma W_{T(t)}$$

being W_T the standard brownian motion.
Log-price time increments over a time-span Δ are hence given by:

$$r_t = v_t - v_{t-\Delta} = \mu(T(t) - T(t-\Delta)) + \sigma\left(W_{T(t)} - W_{T(t-\Delta)}\right)$$

With $r_t \approx N\left(\mu I_t, \sigma^2 I_t\right)$ for $I_t = T(t) - T(t-\Delta)$

As anticipated, the return on a given time span is function of the amount of activity, measured by I_t, that occurred over Δ.
This model is named "mixture of distributions hypothesis" because each return is drawn from a different distribution; it has the advantage, compared to Paretian hypothesis, to work with finite variance, but relaxes the assumption of stationary distribution of returns against empirical fact noticed on long time scales.

1.1.3 An alternative hypothesis: market return distribution is max-stable

The attractiveness of the Paretian hypothesis – remind - comes from its accounting for fat tails having recourse to a stable family of distributions, where the stability is a sort of closure property with respect to the sum of random variables representing the progressive market information updating.
Analogous property is obeyed by distributions of maxima as we know from extreme value theory and is completed with the interesting requirement of finite variance; this circumstance encouraged me in exploring the relationships between market return stochastic processes and extreme value distributions as reported in the remainder of this document.
At the moment, in order to fix preliminary ideas, I introduce only some Extreme Value theory insight that will be developed further.
The fundamental result of EVT is about the asymptotic distribution of the maxima of a sample: like the Central Limit Theorem, which ensures that the sums of random variables are normally distributed, so an important theorem from Gnedenko [38] firstly ensures that maxima necessarily follow a distribution that obeys the max stability property, then shows that only three distributions types are max-stable: Frechet, Gumbel, Weibull.
A non-degenerate distribution G is *max-stable* iff there exist $\{a_n > 0, n \geq 1\}$ and $\{b_n \in \Re, n \geq 1\}$ such that $\left(G\left(a_n x + b_n\right)\right)^n \to G\left(x\right)$
To make explicit the connection between alpha stable distributions and max stable distributions let's focus on *regularly varying* concept.

Definition 1.1.1 *A positive measurable function f is called regularly varying (at infinity) with index $\alpha \in \Re$ if*

1. *It is defined on some neighborhood $[x_0, \infty)$ of infinity*
2. $\forall t > 0 \lim\limits_{x \to \infty} \frac{f(tx)}{f(x)} = t^\alpha$

If $\alpha = 0$ is said to be slowly varying (at infinity)

The regularly varying property characterizes the membership of a distribution to specific domains of attraction here defined:

Definition 1.1.2 *We say that the random variable X and its distribution F belongs to the domain of attraction of the alpha stable distribution G_α if there exist constants $a_n > 0$, $b_n \in \Re$ such that*

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} G_\alpha, n \to \infty$$

holds.

A similar statement defines the maximum domain of attraction:

Definition 1.1.3 *The random variable X and its distribution F belong to the domain of attraction of the max stable distribution H if there exist constants $c_n > 0$, $d_n \in \Re$ such that*
$\frac{M_n - d_n}{c_n} \xrightarrow{d} H$ *as* $n \to \infty$

Now are presented here, without proof, the conditions, based on slowly varying property, to belong respectively to the domain of attraction of an alpha stable law and to maximum domain of attraction of Φ_α.
The distribution F belongs to the domain of attraction of an alpha stable law for some $\alpha < 2$ if and only if:

$$F(-x) = \frac{q + o(1)}{x^\alpha} L(x), \bar{F}(x) = \frac{p + o(1)}{x^\alpha} L(x), x \to \infty$$

where L is slowly varying and p, q are non-negative constants such that $p+q > 0$
The distribution F belongs to the maximum domain of attraction of Φ_α , $\alpha > 0$, if and only if $\bar{F} = x^{-\alpha} L(x)$ for some slowly varying function L.
From these results we can conclude that if $X \in DA(G_\alpha)$ for some alpha stable distribution G_α with $\alpha < 2$ and $P(X > x) \approx cP(|X| > x)$, $c > 0$, as $x \to \infty$ (i.e. this distribution is not totally skewed to the left) then $X \in MDA(\Phi_\alpha)$

Alpha-stable and max-stable distributions are hence intimately connected by the regular variation condition; thanks to this connection a class of max stable distributions can share with alpha stable laws a fundamental closure property directly related to the regular variation. In other words, even the Frecht extreme value distribution is invariant under summation and can be regarded as a valid candidate to represent the market information arrival.
The closure property is as follows:

Theorem 1.1.4 *Let X and Y be two independent, regularly varying, nonnegative random variables with index $\alpha \geq 0$.*
Then $X+Y$ is regularly varying with index $\alpha \geq 0$ and $P(X + Y > x) \sim P(X > x) + P(Y > x)$ as $x \to \infty$

Proof:

$$P(X+Y>x) \geq [P(X>x)+P(Y>x)](1-o(1))$$

from

$$\{X+Y>x\} \supset \{X>x\} \bigcup \{Y>x\}$$

For $0<\delta<\frac{1}{2}$

$$\{X+Y>x\} \subset \{X>(1-\delta)x\} \bigcup \{Y>(1-\delta)x\} \bigcup \{X>\delta x, Y>\delta x\}$$

Implies that

$$P(X+Y>x) \geq [P(X>x)+P(Y>x)](1-o(1))$$

hence:

$$1 \leq \liminf_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)+P(Y>x)} \leq \limsup_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)+P(Y>x)} \leq (1-\delta)^{-\alpha}$$

Letting $\delta \downarrow 0$ the proof is complete.

The need of non gaussian processes for stock returns is theoretically, historically and practically well known and opens promising perspectives when oriented to Extreme Value Theory; what follows would like to be a step in that direction.

Chapter 2

Stochastic integration

The description of stochastic rate interest models is crucial in this research and it is related to an important branch of probability theory concerning the stochastic integration.
In what follows central topics of stochastic integration are recalled so to justify and correctly use a fundamental tool of stochastic calculus, the Ito's rule; Ito's rule is the differentiation rule of stochastic calculus needed to solve Stochastic Differential Equations, and it works quite differently from standard differetiation rules because the special behaviour of Brownian paths.
Brownian paths, as I'm going to show, have infinite variance and, strictly speaking, are not differentiable.
The stochastic integration, hence, can not be intended in a traditional way, and has to be conceived as an approximation in some probabilistic sense; that sense is contextualized now.

Definition 2.0.5
Given a finite set Ω *and the set* F *of all subset in* Ω*, a probability measure* P *is a function from* F *to [0,1] such that:*
i) $P(\emptyset) = 0 \; P(\Omega) = 1$
ii) If $A_1, A_2, ...$ *is a sequence of disjoint sets in* F*, then*

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

Definition 2.0.6 $P(A) = \sum_{\omega \in A} P\{\omega\}$ *for* $A \in F$

Definition 2.0.7 *A* $\sigma-$*algebra is a collection* G *of subsets of* Ω *such that:*

1. $\emptyset \in \mathrm{G}$

2. *If* $\mathrm{A} \in \mathrm{G}$*then its complement* $\mathrm{A}^{\mathrm{C}} \in \mathrm{G}$

3. *If* $A_1, A_2, ...$ *is a sequence of sets in* G *then* $\bigcup_{\mathrm{k}=1}^{\infty} A_k$ *is also in* G

Definition 2.0.8 *A filtration is a sequence of* $\sigma-$*algebras* $F_0, F_1, ..., F_n$ *such that each of them contains all the sets belonging to the previous* $\sigma-$*algebra*

Definition 2.0.9 *A random variable is a function mapping* Ω *into* $\Re$

Definition 2.0.10 *Given* F*, the* $\sigma-$ *algebra of all subset of* Ω*, and* X *a random variable on* (Ω, F) *, the* $\sigma-$*algebra* $A \subset \Re$ *generated by* X *is the collection of all sets* $\sigma(X)$*, with* $\{\omega \in \Omega; X(\Omega) \in A\}$*.*

Being G *a sub -*$\sigma-$ *algebra of* F *,* X *is G-measurabile if every set in* $\sigma(X)$ *is also in* G

Definition 2.0.11 *A stochastic process is a parametrized collection of random variables* $\{X_t\}_{t \in t}$
defined on a probability space (Ω, F, P) *and assuming values in* $(\Re^n)$

Definition 2.0.12 *For any set* $A \subset \Re$ *the induced measure of* A *is:* $L_X(A) \triangleq P\{X \in A\}$

Definition 2.0.13 *Given a probability space* (Ω, F, P) *a brownian motion* $B(t, \omega) : [0, \infty) \times \Omega \to \Re$*is a stochastic process with the following properties:*

1. $P\{\omega; B(0, \omega) = 0\} = 1$

2. $B(t)$ *is a continuous function of* t

3. *if* $0 \leq k \leq n$ *then* $\forall k$ *the increments* $B(t_k) - B(t_{k-1}) \cong N(0, t_k - t_{k-1})$ *are independent*

2.1 The Stochastic Integral

To introduce the stochastic integration it is useful to recognize the circumstances that claim for this mathematical tool and give some evidence about the problems arising from the brownian motion integration and the solution found to circumvent it.

Suppose to have an amount X invested in risky assets and that the assets variation is driven by the following process:

$$\Delta X = \mu X \Delta t + \sigma X \Delta B$$

In a short time step $[t, t + \Delta t]$a part of the asset variation is proportional to the asset amount and to the time step, another part is proportional to the random quantity $\sigma \Delta B$ that represents the stochastic effect of combined market factors.

To evaluate the amount of X(t) at the T instant, subdivide the time lapse [0,T] in N short intervals Δt, and sum the increments (or decrements) of X occurred in each of the intervals; hence, for $\Delta t = \frac{T}{N}$ it is:

$$X(T) - X(0) = \sum_{i=1}^{N} X(t_{i-1})(t_i - t_{i-1}) + \sum_{i=1}^{n-1} \sigma X(t_{i-1})(B(t_i) - B(t_{i-1}))$$

with $\forall$ i = 0, 1, ..., N. $t_i = i\frac{T}{N}$

If the limit for $N \to +\infty$ exists, the above sums can be written in integral form:

$$X(T) - X(0) = \int_0^T \mu X(t)\, dt + \int_0^t \sigma X(t)\, dB(t)$$

A closer look to this equation shows that the first integral is well defined but the second one, being the integrand a stochastic function, cannot be interpreted in the usual way because the limit of the sums

$$\sum_{i=1}^{n-1} \sigma X(t_{i-1})(B(t_i) - B(t_{i-1}))$$

loose its traditional meaning; intuitively it is clear that, having the Brownian motion unbounded variations, it is not differentiable as will be proved shortly.

Starting from this fact it becomes fundamental to describe the conditions to ensure the existence of the stochastic integral; it will be the case after preliminary considerations.

2.1.1 Quadratic variation and non-differentiability of brownian motion

To give adequate evidence to the non-differentiability of Brownian motion we focus on its quadratic variation; the result obtained here will be an essential part of the Ito's Lemma proof.

Theorem 2.1.1 *Let B(t) be a brownian motion ; in L2:*

$$\int_a^b \left[B\left(t_i\right) - B\left(t_{i-1}\right)\right]^2 = (b-a)$$

$$[a,b] \subset [0,T]\, \forall b, \forall a \qquad t.c. \qquad b \geq a$$

Proof:

Fixed an arbitrary division $\{t_0, t_1, ..., t_n\}$ of $[a,b]$

$E\left[\sum_{t=1}^n \left[\left(B\left(t_i\right) - B\left(t_{i-1}\right)\right)^2\right]\right] = \sum_{i=1}^n E\left[\left(B\left(t_i\right) - B\left(t_{i-1}\right)\right)^2\right]$
$= \sum_{i=1}^n E\left[B\left(t_i - t_{i-1}\right)^2\right] = \left(t_i - t_{i-1}\right)$

$Var \sum_{i=0}^n \left[\left(B\left(t_i\right) - B\left(t_{i-1}\right)\right)^2\right] =$
$= \sum_{i=1}^n Var\left[\left(B\left(t_i\right) - B\left(t_{i-1}\right)\right)^2\right] = \sum_{i=1}^n Var\left[B\left(t_i - t_{i-1}\right)^2\right]$
$= \sum_{i=1}^n \left\{E\left[B\left(t_i - t_{i-1}\right)^4\right] - E^2\left[B\left(t_i - t_{i-1}\right)^2\right]\right\}$
$= \sum_{i=1}^n \left\{3\left(t_i - t_{i-1}\right)^2 - \left(t_i - t_{i-1}\right)^2\right\}$
from the known relation
$\left(E\left[X^4\right] = 3Var\left[X^2\right]\right)$

Futhermore:

$$\sum_{i=1}^n \left(t_i - t_{i-1}\right)^2 \leq \max\left(t_i - t_{i-1}\right) \sum_{i=1}^n \left(t_i - t_{i-1}\right)$$
$$\lim_{\max(t_i - t_{i-1}) \to 0} \max\left(t_i - t_{i-1}\right) \sum_{i=1}^n \left(t_i - t_{i-1}\right) = 0$$

i.e. the variance of $\sum_{i=0}^n \left[\left(B\left(t_i\right) - B\left(t_{i-1}\right)\right)^2\right]$ tends to zero.
The statement: $\lim_{\Delta t \to 0} \sum_{i=1}^n \left[B\left(t_i\right) - B\left(t_{i-1}\right)\right]^2 = \left(t_t - t_{i-1}\right)$, is confirmed.

This result implies the non-differentiability of the brownian motion; if it would be differentiable, through the application of the average value theorem, we would obtain:

$\sum_{i=1}^n \left(B\left(t_i\right) - B\left(t_{i-1}\right)\right)^2 = \sum_{i=1}^n \left|f'\left(t*_{i-1}\right)\right|^2 \left(t_i - t_{i-1}\right)^2$
$\leq \max\left(t_i - t_{i-1}\right) \sum_{i=1}^n \left|f'\left(t*_{i-1}\right)\right|^2 \left(t_i - t_{i-1}\right)$

$$\lim_{\max(t_i - t_{i-1}) \to 0} \sum_{i=1}^{n} \left(B\left(t_i\right) - B\left(t_{i-1}\right) \right)^2$$
$$\leq \lim_{\max(t_i - t_{i-1}) \to 0} \left[\max\left(t_i - t_{i-1}\right) \sum_{i=1}^{n} \left| f'\left(t*_{i-1}\right) \right|^2 \left(t_i - t_{i-1}\right) \right] =$$
$$= \lim_{\max(t_i - t_{i-1}) \to 0} \left[\max\left(t_i - t_{i-1}\right) \right] \int_0^t \left| f'\left(t*_{i-1}\right) \right|^2 dt = 0$$

clearly different from the previous result.

Moreover, integrating the Brownian motion with the rules of standard calculus, the outcome:

$$\int_a^b B\left(t\right) dB\left(t\right) = \frac{1}{2} \left(B\left(b\right)^2 - B\left(a\right)^2 \right)$$

is incorrect.

Observe that:

$$B^2\left(b\right) - B^2\left(a\right) = \sum_{i=1}^{n} \left(B^2\left(t_i\right) - B^2\left(t_{i-1}\right) \right)$$
$$= \sum_{i=1}^{n} \left(B\left(t_i\right) - B\left(t_{i-1}\right) \right)^2 + 2 \sum_{i=1}^{n} B\left(t_i\right) B\left(t_{i-1}\right) - 2 \sum_{i=1}^{n} B\left(t_{i-1}\right)^2$$
$$= \sum_{i=1}^{n} \left(B^2\left(t_i\right) - B^2\left(t_{i-1}\right) \right)^2 + 2 \sum_{i=1}^{n} B\left(t_{i-1}\right) \left(B\left(t_i\right) - B\left(t_{i-1}\right) \right)$$

It follows that:

$$\sum_{i=1}^{n} B\left(t_{i-1}\right) \left(B\left(t_i\right) - B\left(t_{i-1}\right) \right) = \frac{1}{2} \left(B^2\left(b\right) - B^2\left(a\right) \right) - \frac{1}{2} \sum_{i=1}^{n} \left(B\left(t_i\right) - B\left(t_{i-1}\right) \right)^2$$

$$= \lim_{\Delta t \to 0} \sum_{i=1}^{n} B\left(t_{i-1}\right) \left(B\left(t_i\right) - B\left(t_{i-1}\right) \right) = \frac{1}{2} \left(B^2\left(b\right) - B^2\left(a\right) \right) - \frac{1}{2} \left(b - a\right)$$

because $\lim_{\Delta t \to 0} \sum_{i=1}^{n} \left[B\left(t_i\right) - B\left(t_{i-1}\right) \right]^2 = \left(t_t - t_{i-1}\right)$

This result differs from the result obtained from the standard calculus for the term $-\frac{1}{2}\left(b - a\right)$.

2.1.2 The Ito integral

Due to the non-differentiability of Brownian motion, its integration has to be defined out of the rules of the ordinary calculus adopting a strategy that firstly describes an "elementary integrand" and then states a sequence of elementary processes that approximates the general process.
This section shows the construction of the Ito integral and its relevant properties. Given the Brownian motion $B\left(t\right), t \geq 0$ with the following properties:

1. $s \leq t \Rightarrow F\left(s\right) \subseteq F\left(t\right)$
2. $\forall t B\left(t\right)$ is $F\left(t\right) - measurable$
3. For $t_0, t_1, ..., t_n$ the increments $B\left(t_1\right) - B\left(t_0\right), B\left(t_2\right) - B\left(t_1\right), ..., B\left(t_n\right) - B\left(t_{n-1}\right)$ are independent

and a function $\delta(t), t \geq 0$ where:

1. $\forall t \delta(t)$ is $F(t) - measurable$

2. $\forall t E \int_o^T \delta^2(t)\, dt < \infty$ i.e. is *square-integrable*

define the Ito integral as

$$I(t) = \int_o^t \delta(u)\, dB(u)\, t \geq 0$$

If, given a partition of $[0,T]$ such that $0 = t_0 \leq t_1, ..., \leq t_n = T$, $\delta(t) = \delta(t_k)$ $t_{i-1} \leq t_k < t_i$ ($\delta(t)$ is constant for each sub interval $t_{i-1} \leq t_k < t_i$) then $\delta(t)$ is called an *elementary process*

Remark that the Ito integral for an elementary process is a martingale

Theorem 2.1.2 *Consider:*

$$I(t) = \sum_{j=0}^{k-1} \delta(t_j)\left[B(t_{j+1}) - B(t_j)\right] + \delta(t_k)\left[B(t) - B(t_k)\right]$$

$$t_k \leq t \leq t_{k+1}$$

consider also partition points t_l and $t_k$such that $s \in [t_l, t_{l+1}]$ and $t \in [t_k, t_{k+1}]$ with $0 \leq s \leq t$

$I(t_{k+1})$ is a martingale: $E\left[I(t_{k+1}) | F(s)\right] = I(s)$

Proof:

$I(t) = \sum_{j=0}^{l-1} \delta(t_j)[B(t_{j+1}) - B(t_j)] + \delta(t_l)[B(t_{l+1}) - B(t_l)] +$
$+ \sum_{j=l+1}^{k-1} \delta(t_j)[B(t_{j+1}) - B(t_j)] + \delta(t_k)[B(t) - B(t_k)]$

Conditional expectation for the first two terms is:

$$E\left[\sum_{j=0}^{l-1} \delta(t_j)[B(t_{j+1}) - B(t_j)] \middle| F(s)\right] = \sum_{j=0}^{l-1} \delta(t_j)[B(t_{j+1}) - B(t_j)]$$

$E[\delta(t_l)[B(t_{l+1}) - B(t_l)]|F(s)] = \delta(t_l)[E[B(t_{l+1})|F(s)] - B(t_l)]$
$= \delta(t_l)[B(s) - B(t_l)]$

the remaining terms return null value:

$E\left[\sum_{j=l+1}^{k-1} \delta(t_j)[B(t_{j+1}) - B(t_j)] \middle| F(s)\right]$
$= \sum_{j=l+1}^{k-1} E[[E[\delta(t_j)[B(t_{j+1}) - B(t_j)]]|F(t_j)]|F(s)]$
$= \sum_{j=l+1}^{k-1} [E[\delta(t_j)[E[B(t_{j+1})|F(t_j)] - B(t_j)]]|F(s)] = 0$

$E[\delta(t_k)[B(t) - B(t_k)]|F(s)] = E[\delta(t_k)[E[B(t)|F(t_k)] - B(t_k)]|F(s)] = 0$

Another relevant property of Ito integral is the Ito isometry:

Theorem 2.1.3

$$E\left[I^2(t)\right] = E\left[\int_0^t \delta^2(u)\,du\right]$$

Proof

$I^2(t) = \left[\sum_{j=0}^{k} \delta(t_j)[B(t_{j+1}) - B(t_j)]\right]^2$
$= \sum_{j=0}^{k} \delta^2(t_j)[B(t_{j+1}) - B(t_j)]^2 + \sum_{i<j} \delta(t_i)\delta(t_j)[B(t_{i+1}) - B(t_i)][B(t_{j+1}) - B(t_j)]$

The cross term average is null because the independence of the Brownian motions.

$E\left[I^2(t)\right] = \sum_{j=0}^{k} E\left[\delta^2(t_j)[B(t_{j+1}) - B(t_j)]^2\right]$
$= \sum_{j=0}^{k} E\left[\delta^2(t_j) E\left[[B(t_{j+1}) - B(t_j)]^2 \middle| F(t_j)\right]\right]$
$= \sum_{j=0}^{k} E\left[\delta^2(t_j)(t_{j+1} - t_j)\right]$
$= E\sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \delta^2(u)\,du$
$= E\left[\int_0^t \delta^2(u)\,du\right]$

The construction of the Ito's integral, once defined the elementary integrand, comes from the approximation to it, in mean square limit, of the general integrand.

The existence of the Ito's integral is hence proofed by the existence of such a limit:

Theorem 2.1.4 *Given a function* $\delta(t), t \geq 0$ *where:*

1. $\forall t\ \delta(t)$ *is* $F(t) - measurable$

2. $\forall t\ E\left[\int_o^T \delta^2(t)\, dt\right] < \infty$ *i.e. is square-integrable*

there is a sequence of elementary processes $\{\delta_n\}_{n=1}^{\infty}$ *such that*

$$\lim_{n\to\infty} E\int_0^T [\delta_n(t) - \delta(t)]^2\, dt = 0$$

Proof:
Suppose n and m large positive integers

$$\begin{array}{l} var\left(I_n(t) - I_m(t)\right) = E\left(\int_0^T [\delta_n(t) - \delta_m(t)]\, dB(t)\right)^2 \\ \underset{n\to\infty}{} \\ = E\int_0^T [\delta_n(t) - \delta(t)_m]^2\, dt \\ = E\int_0^T [[\delta_n(t) - \delta(t)] + [\delta(t) - \delta_m(t)]]^2\, dt \\ \leq 2E\int_0^T [\delta_n(t) - \delta(t)]^2\, dt + 2E\int_0^T [\delta_m(t) - \delta(t)]^2\, dt \end{array}$$

being

$$(a+b)^2 \leq 2a^2 + 2b^2$$

2.2 Ito's formula

Having the tools to extend integration to stochastic processes, it is possible to develop the operational set needed to solve Stochastic Differential Equations, i.e. equations whose unknown function is a stochastic process.

Consider the Ito process:

$$dY(t) = \mu Y(t)\, dt + \sigma Y(t)\, dB(t)$$

and a function $f = f(t,x) = f(t, Y(t))$ written on it. Suppose we have to calculate its derivative; by Taylor's rule:

$$df = \frac{\partial f}{\partial x} dY(t) + \frac{\partial f}{\partial t} dt + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} dY(t)^2 + \frac{\partial^2 f}{\partial x \partial t} dY(t)\, dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2} dt^2$$

Adequately replacing the $dY(t)$ terms:

$$\frac{\partial f}{\partial x} dY(t) = \frac{\partial f}{\partial x}\left[\mu Y(t)\, dt + \sigma Y(t)\, dB(t)\right] \frac{1}{2}$$

$$\frac{\partial^2 f}{\partial x^2} dY(t)^2 = \frac{\partial^2 f}{\partial x^2}\left(\frac{1}{2}\mu^2 Y(t)^2\, dt^2 - \mu^2 \sigma^2 Y(t)^2\, dt dB(t) + \frac{1}{2}\sigma^2 Y(t)^2 \left[dB(t)\right]^2\right)$$

Observing that:

1. $(dt)^2$ or $(dt)(dB(t))$ terms are negligible

2. $(dB(t))^2 \cong dt$

we obtain:

$$df = \left(\frac{\partial f}{\partial x}\mu Y(t) + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}\sigma^2 Y(t)^2\right) dt + \frac{\partial f}{\partial x}\sigma Y(t)\, dB(t)$$

2.3 Feynman-Kac formula

Some classes of solution of PDE have a representation in terms of expected value; this representation, as well as to be useful for practical purposes, provides as an important bridge between probability and calculus.

Theorem 2.3.1 *Given the PDE:*

$$\frac{\partial f}{\partial t} + \mu(x,t)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 f}{\partial x^2} = 0$$
$$f(x,T) = \psi(x)$$

it is: $f(x,T) = E[\psi(X_T)|X_t = x]$
with $dx(t) = \mu(x,t)dt + \sigma(x,t)dW_t$

proof:
From Ito's lemma:

$$df = \left(\mu(x,t)\frac{\partial f}{\partial x} + \frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma(x,t)\frac{\partial f}{\partial x}dW_t$$

$$\int_t^T df = f(X_T,T) - f(x,t) = \int_t^T \sigma(x,t)\frac{\partial f}{\partial x}dW_t$$

Re-arranging and taking the expected value:

$$f(x,t) = E[f(X_T,T)] - E\left[\int_t^T \sigma(x,t)\frac{\partial f}{\partial x}dW_t\right]$$

but: $E\left[\int_t^T \sigma(x,t)\frac{\partial f}{\partial x}dW_t\right] = 0$
hence: $f(x,t) = E[f(X_T,T)] = E[\psi(X_T)] = E[\psi(X_T)|X_t = x]$

Chapter 3

Interest rates modeling – the Vasicek seminal approach

In 1977 Vasicek wrote one of the earliest papers known about the model for the risk-free rate of interest [82]; in the same paper, Vasicek also developed a more general approach to pricing which ties in with what we now refer to the as the risk-neutral-pricing approach, and this fact has, from a methodological point of view, the same relevance of the main result itself.
For notational convenience and clarity of exposition, we will restrict ourselves here to one-factor, diffusion models for the short rate, $r(t)$. The approach is, however, easily extended to multifactor models.

The class of unifactorial models supplies a representation of the term structure of interest rates.
Formally the term structure - or Yield Curve - is a function of the *simply compounded spot interest rates* ($L(t,T)$) and of the *annually compounded spot interest rates* ($Y(t,T)$) , so defined:

Definition 3.0.2 .

Let

$\tau(t,T)$ *be a year fraction between* t *and* T
$r(t,T)$ *be the value at time* t *of interest rates for a bond with maturity* T
$P(t,T)$ *be the price at time* t *of a bond with maturity* T

Then

$$L(t,T) = \frac{1 - P(t,T)}{\tau(t,T) P(t,T)}$$

$$Y(t,T) = \frac{1 - P(t,T)}{P(t,T)^{1/\tau(t,T)}} - 1$$

Hence the term structure is the curve of the function:

$$T \mapsto \begin{cases} L(t,T) & t < T \leq t \\ Y(t,T) & T > t+1 \end{cases}$$

In unifactorial models the term structure is driven by a single variable, namely the spot rate; the spot rate evolves in time according to a diffusive process:

$$dr(t) = \alpha(r_t, t)\, dt + \beta(r_t, t)\, dB$$

where B is a standard Brownian motion and $\alpha(r_t, t)$ and $\beta(r_t, t)$ are characteristic for each distinct model in the class.
Specific models for $r(t)$ include Vasicek:

$$dr(t) = \alpha(\mu - r(t))\, dt + \sigma d\tilde{B}$$

and Cox, Ingersoll and Ross [19]:

$$dr(t) = \alpha(\mu - r(t))\, dt + \sigma\sqrt{r(t)} d\tilde{B}$$

These processes belong to the set of "mean reverting model": at each time step the interest rate value is updated through a partial subtraction of the current amount of the interest rate value itself, so that the new amount tends to an equilibrium point μ; this behavior appears to be consistent with interest rates real market fluctuations and explains the success of Vasicek and CIR formulations in literature and in professional practice.
Both models give rise to analytical formulae for zero-coupon bond prices and European options of the same bonds. The CIR model has the advantage that interest rates stay positive because of the square-root of $r(t)$ in the volatility.
Both are also examples of affine term-structure models: that is, the solution $D(t,T)$ of $dr(t)$ can be written in the form

$$D(t,T) = \exp(A(T-t) - B(T-t) r(t))$$

for suitable functions A and B.

3.1 Non arbitrage models

The Vasicek and CIR models are examples of time-homogeneous, equilibrium models. A disadvantage of such models is that they give a set of theoretical prices for bonds which will not normally match precisely the actual prices that we observe in the market. This led to the development of some time-inhomogeneous Markov models for $r(t)$, most notably those due to Ho & Lee [44]

$$dr(t) = \phi(t)\,dt + \sigma d\tilde{B}$$

Hull & White [46]

$$dr(t) = \alpha(\mu - r(t))\,dt + \sigma(t)\,d\tilde{B}$$

and Black & Karasinski [9]

$$dr(t) = \alpha(t)\,r(t)\,[\theta(t) - \log r(t)]\,dt + \sigma(t)\,r(t)\,d\tilde{B}$$

In each of these models all deterministic functions of t are calibrated in a way which gives a precise match at the start date (say time 0) between theoretical and observed prices of zero-coupon bonds (Ho & Lee) and possibly also some derivative prices.
For example, at-the-money interest-rate caplet prices could be used to derive the volatility function, in the Hull & White model.
Because these models involve an initial calibration of the model to observed prices there is no arbitrage opportunity at the outset. Consequently these models are often described as no-arbitrage models. In contrast, the time-homogeneous models described earlier tell us that if the prices were to evolve in a particular way then the dynamics will be arbitrage free.
The Ho & Lee and Hull & White models are also examples of affine term-structure models. The Black & Karasinski model does not yield any analytical solutions, other than that $r(t)$ is log-normally distributed. However, the BK model is amenable to the development of straightforward and fast numerical methods for both calibration of parameters and calculation of prices.
It is standard market practice to recalibrate the parameters and time-dependent, deterministic functions in these no-arbitrage models on a frequent basis. For example, take two distinct times $T_1 < T_2$. In the Hull & White model we would calibrate it at time T_1 for all $t < T_1$ to market prices; at time T_2 we would repeat this calibration using prices at T_2 resulting in the model for $t > T_2$. If the Hull & White model is correct then we should find the congruence between the outcomes of the model calibrated at time T_1 and the values used to re-calibrate the same model in T_2; in practice this rarely happens so that we end up treating model parameters as stochastic rather than the deterministic form assumed in the model.
In 1992 Heath, Jarrow & Morton [42] proposed a framework that represents a substantial leap forward in how the term structure of interest rates is perceived and modeled.
Previously models concentrated on modeling of $r(t)$ and other relevant quantities in a multifactor model; the HJM framework instead of focusing on $r(t)$

model the instantaneous forward-rate curve, $f(t;T)$, directly. Given the forward-rate curve we then immediately get:

$$D(t,T) = \exp\left[-\int_t^T f(t,u)\right] du$$

Let's follow the derivation of HJM framework from an Ito process defined as usual:

$$dp(t,T) = r(t)\, p(t,T)\, dt + v(t,T,\Omega)\, p(t,T)\, dB$$

with:

$p(t,T)$: Price at time t of a zero-coupon bond with principal 1 maturing at time T

Ω_t: Vector of past and present values of interest rates and bond prices at time t

$v(t,T,\Omega)$: Volatility of $P(t,T)$

$f(t,T_1T_2)$: Forward rate as observed at time t for the period between time T1 and time T2

$r_{(t)}$: Short term risk free interest rate at time t (with $P_{(t_1,t_2)} = e^{r_{(t_1,t_2)}(t_2-t_1)}$)

dB: Brownian motion

Apply the Ito's lemma:

$$\frac{\partial \ln(p(t,T))}{\partial p(t,T)} r(t)\, P(t,T)\, dt = \frac{1}{P(t,T)} r(t)\, P(t,T)\, dt = r(t)\, dt$$

$$\frac{\partial \ln(p(t,T))}{\partial t} dt = 0$$

$$\frac{\partial \ln(p(t,T))}{\partial p(t,T)} v(t,T,\Omega)\, P(t,T)\, dB = \frac{1}{P(t,T)} v(t,T,\Omega)\, P(t,T)\, dB = v(t,T,\Omega)\, dB$$

$$\frac{\partial^2 \ln(p(t,T))}{2\partial p^2(t,T)} v^2(t,T,\Omega)\, P^2(t,T)\, dt = -\frac{1}{P^2(t,T)} v^2(t,T,\Omega)\, P^2(t,T)\, dt = -\frac{v^2(t,T,\Omega)}{2} dt$$

joining the parts:

$$d\ln[p(t,T)] = \left[r(t) - \frac{v^2(t,T,\Omega)}{2}\right] dt + v(t,T,\Omega)\, dB(t) \qquad (3.1)$$

In order to avoid arbitrage opportunities an investment from t to T_2 must have the same return as two compounded investments from t to T_1 and from T_1 to T_2 :

$$r_{(T_1,T_2)} = \frac{r_{(t,T_2)}(T_2) - r_{(t,T_1)}(T_1 - t)}{(T_2 - T_1)}$$

From the definition of $P_{(t_1,t_2)}$ it is:

$$r_{(t_1,t_2)}(t_2 - t_1) = \ln\left[P_{(t_1,t_2)}\right]$$

and hence:

$$f(t, T_1T_2) = \frac{\ln[p(t,T_1)] - \ln[p(t,T_2)]}{T_2 - T_1}$$

From (3.1) we have:

$$df(t,T_1,T_2) = \left[\frac{v(t,T_2,\Omega)^2 - v(t,T_1,\Omega)^2}{2(T_2 - T_1)}\right]dt + \left[\frac{v(t,T_1,\Omega) - v(t,T_2,\Omega)}{(T_2 - T_1)}\right]dB$$

3.2 Deriving the interest rates PDE

The price of a contract written on interest rates can be derived from the absence of arbitrage principle that allows the composition of a riskless portfolio Π in an infinitesimal timestep $[t, t + dt]$; the absence of arbitrage principle states the impossibility to have gain without risk ("*no free lunch principle*"): all available investment opportunities that compensate (immunize) their risks have identical value because – given a set of contracts sharing the same risk level - nobody accepts to buy less profitable ones.
Let $V_1(t, r_t)$e $V_2(t, r_t)$ be the values of two contracts assembled in a portfolio Π.
On the basis of Ito's formula the variation of $V_i(t, r_t)$ (i = 1, 2) is given by:

$$dV_i(t,r_t) = \left[\frac{\partial V_i(t,r_t)}{\partial t} + \alpha(r_t)\frac{\partial V_i(t,r_t)}{\partial r} + \frac{1}{2}\beta^2(r_t)\frac{\partial^2 V_i(t,r_t)}{\partial r^2}\right]dt + \left[\beta(r_t,t)\frac{\partial V_i(t,r_t)}{\partial r}\right]dB$$

put:

$$\mu_t(t,r_t)V_i(t,r_t) = \frac{\partial V_i(t,r_t)}{\partial t} + \alpha(r_t)\frac{\partial V_i(t,r_t)}{\partial r} + \frac{1}{2}\beta^2(r_t)\frac{\partial^2 V_i(t,r_t)}{\partial r^2}$$

$$\sigma_t(t,r_t)V_i(t,r_t) = \beta(r_t,t)\frac{\partial V_i(t,r_t)}{\partial r}$$

it is:

$$dV_i(t,r_t) = \mu_t(t,r_t)V_i(t,r_t) + \sigma_t(t,r_t)V_i(t,r_t)\,dB$$

Each contract gives contribution q_i to the portfolio Π;Π variations are consequently:

$$\begin{aligned} d\Pi(t,r_t) = {} & [q_1\mu_1(t,r_t)V_1(t,r_t) + q_2\mu_2(t,r_t)V_2(t,r_t)]\,dt + \\ & + [q_1\sigma_1(t,r_t)V_1(t,r_t) + q_2\sigma_2(t,r_t)V_2(t,r_t)]\,dB \end{aligned}$$

Being Π riskless its value increases or decrease at risk free interest rate:

$$d\Pi(t,r_t) = r(t)\Pi(t,r_t)\,dt = r(t)[q_1V_1(t,r_t) + q_2V_2(t,r_t)]\,dt$$

or, equivalently :

$$\begin{aligned}[q_1\mu_1(t,r_t)V_1(t,r_t)+q_2\mu_2(t,r_t)V_2(t,r_t)] &= r(t)[q_1V_1(t,r_t)+q_2V_2(t,r_t)]\\ [q_1\sigma_1(t,r_t)V_1(t,r_t)+q_2\sigma_2(t,r_t)V_2(t,r_t)] &= 0\end{aligned}$$

$$\begin{aligned}q_1V_1(t,r_t)(\mu_1(t,r_t)-r(t))+q_2V_2(t,r_t)(\mu_2(t,r_t)-r(t)) &= 0\\ [q_1\sigma_1(t,r_t)V_1(t,r_t)+q_2\sigma_2(t,r_t)V_2(t,r_t)] &= 0\end{aligned}$$

from second equation:

$$q_2V_2(t,r_t) = -q_1V_1(t,r_t)\frac{\sigma_1(t,r_t)}{\sigma_2(t,r_t)}$$

after substitution in first equation:

$$q_1V_1(t,r_t)(\mu_1(t,r_t)-r(t)) - q_1V_1(t,r_t)\frac{\sigma_1(t,r_t)}{\sigma_2(t,r_t)}(\mu_2(t,r_t)-r(t)) = 0$$

$$\frac{(\mu_1(t,r_t)-r(t))}{\sigma_1(t,r_t)} = \frac{(\mu_2(t,r_t)-r(t))}{\sigma_2(t,r_t)}$$

The function:

$$\lambda(t,r_t) = \frac{(\mu(t,r_t)-r(t))}{\sigma(t,r_t)}$$

is named "price for risk", from which:

$$\mu(t,r_t) = r(t)+\lambda(t,r_t)\sigma(t,r_t)$$

Replacing $\mu(t,r_t)$ and $\sigma(t,r_t)$with their explicit forms it becomes:

$$\frac{1}{V}\left[\frac{\partial V_i(t,r_t)}{\partial t}+\alpha(r_t)\frac{\partial V_i(t,r_t)}{\partial r}+\frac{1}{2}\beta^2(r_t)\frac{\partial^2 V_i(t,r_t)}{\partial r^2}\right] = r(t)+\lambda(t,r_t)\left[\frac{1}{V}\beta(r_t)\frac{\partial V_i(t,r_t)}{\partial r}\right]$$

$$\frac{\partial V_i(t,r_t)}{\partial t}+(\alpha(r_t)-\beta(r_t)\lambda(t,r_t))\frac{\partial V_i(t,r_i)}{\partial r}+\frac{1}{2}\beta^2(r_t)\frac{\partial^2 V_i(t,r_t)}{\partial r^2}-r(t)V_i(t,r_t) = 0 \tag{3.2}$$

The solution of this PDE with condition $V(T,r_t)$ returns the price of a bond expiring in $t=T$.
This approach is very instructive from a conceptual point of view, because, mimicking the strategy that Black & Scholes adopted to achieve the option pricing PDE, transforms a stochastic differential equation to a partial differential equation providing a "risk immunization" argument that drops out the stochastic term.

3.3 Vasicek SDE solution

The process satisfying (3.2) can be obtained from the solution to a general SDE known as Ornstein-Uhlenbeck SDE:

$$dX_t = a(t) X_t dt + \sigma(t) dB_t$$

with $X(0) = X_0$
Suppose: $X_t = e^{A(t)} Y_t$
where $A(t)$ is an opportune deterministic function and Y_t is an opportune process whose differential is:

$$dY_t = f(t) dt + \phi(t) dB_t$$

Now the task is to determine $A(t)$, $f(t)$, $\phi(t)$ so that $X_t = e^{A(t)} Y_t$ satisfies the Ornstein-Uhlenbeck equation.
Deriving $X_t = e^{A(t)} Y_t$ it is:

$$dX_t = A'(t) e^{A(t)} Y(t) dt + e^{A(t)} dY(t) = A'(t) X_t dt + e^{A(t)} f(t) dt + e^{A(t)} \phi(t) dB_t$$

from which:

$$e^{A(t)} \phi(t) = \sigma(t), A'(t) = a(t), f(t) = 0$$

and then:

$$\phi(t) = \sigma(t) e^{-\int_0^t a(s) ds}$$

$$Y_t = \int_0^t \sigma(\tau) e^{-\int_0^{\tau t} a(s) ds} dB_\tau + Y_0$$

$$X_t = e^{\int_0^t a(s) ds} \left(\int_0^t \sigma(\tau) e^{-\int_0^{\tau t} a(s) ds} dB_\tau + X_0 \right)$$

This is the general solution of the Ornstein-Uhlenbeck equation.
Turning back to the specific Vasicek SDE:

$$dX_t = \theta(\mu - X_t) dt + \sigma dB_t$$

through the substitution:

$$X_t^* = (\mu - X_t)$$

it becomes

$$dX_t^* = \theta X_t^* dt + \sigma dB_t$$

with:

$$a(t) = -\theta$$

Applying the general solution just found:

$$X_t^* = X_0 e^{-\theta(t-0)} + e^{-\theta t} \sigma \int_0^t e^{-\theta \tau} dB_\tau$$

$$X_t = \mu + (X_0 - \mu) e^{-\theta(t-0)} + e^{-\theta t} \sigma \int_0^t e^{-\theta \tau} dB_\tau$$
$$X_t = \mu \left(1 - e^{-\theta(t-0)}\right) + X_0 e^{-\theta(t-0)} + e^{-\theta t} \sigma \int_0^t e^{-\theta \tau} dB_\tau$$

3.4 Vasicek model calibration

As previously mentioned, Vasicek model is the tool that I'll use in chapter 5 to illustrate my analysis on extreme interest rates values; I will discuss here some technical aspect regarding the estimation of model parameters and the fitting of the model to real market data (model calibration).
To calibrate the Vasicek model the discrete form of the process is used:

$$x(t_i) = c + bx(t_{i-1}) + \delta\varepsilon(t_i)$$
$$c = \mu\left(1 - e^{-\theta\Delta t}\right)$$
$$b = e^{-\theta\Delta t}$$
$$\varepsilon \approx N(0,1)$$

The volatility of the innovation can be deduced by the Ito isometry:

$$E\left[\sigma\int_0^t e^{-\theta\tau}dB_\tau\right]^2 = \sigma^2\int_0^t e^{-2\theta\tau}d\tau = \sigma^2\left(1 - e^{-2\theta t}\right)/2\theta$$
$$\delta = \sigma\sqrt{\left(1 - e^{-2\theta\Delta t}\right)/2\theta}$$

The calibration process is simply an OLS regression of the time series $x(t_i)$ on its lagged form $x(t_{i-1})$. The OLS regression provides the maximum likelihood estimator for the parameters c, b and d. By resolving the three equations system one gets the following α, θ and σ parameters:

$$\begin{aligned}\alpha &= -\ln(b)/\Delta t\\ \theta &= c/(1-b)\\ \sigma &= \delta\Big/\sqrt{\left(b^2-1\right)\Delta t/2\ln(b)}\end{aligned} \tag{3.3}$$

$$\hat{b} = \frac{n\sum_{t-1}^n x_t x_{t-1} - \sum_{t-1}^n x_t \sum_{t-1}^n x_{t-1}}{\sum_{t-1}^n x_{t-1}^2 - \left(\sum_{t-1}^n x_{t-1}\right)^2}$$

$$\hat{c} = \frac{\sum_{t-1}^n \left(x_t - \hat{b}x_{t-1}\right)}{n\left(1-\hat{b}\right)}$$

$$\hat{\delta}^2 = \frac{1}{n}\sum_{t=1}^n \left[x_i - \hat{b}x_{i-1} - \hat{\theta}\left(1-\hat{b}\right)\right]^2$$

From these estimators, using (3.3) one obtains immediately the estimators for the parameters α, θ and σ.

Chapter 4

Extreme Value Theory

The statistical study of maxima plays a central role in this research as declared introducing its main topic: distributional assumptions of classical quantitative finance are dramatically violated during the current economic crisis, and a structured knowledge about the fundamentals of extreme value theory is a solid support in applying its results.

EVT focuses on statistical behavior of $M_n = \max\{a_1, ..., a_n\}$; in theory the distribution of M_n can be derived exactly:

$$\begin{aligned} \Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, ..., X_n \leq z\} \\ &= \Pr\{X_1 \leq z\} \times ... \times \Pr\{X_n \leq z\} \\ &= \{F(z)\}^n \end{aligned}$$

In practice this is not helpful since F is unknown; one possibility is estimate F from observed data and then substitute the estimation in $F(z)^n$ but small discrepancies in estimation of F lead to substantial discrepancies for $F(z)^n$

Starting from this fact the pioneers of EVT field, began to look for approximate families of models for $F(z)^n$ which can be estimated on the basis of extreme data only; this mission involved, from the twenties of the past century, the autonomous contributions from top scientists like Fisher [32], Frechet [34], Gnedenko [38], Gumbel [39], and reached a result that can be thought as the analog of central limit theorem for the maxima (instead of sums) of a series: the explicit asymptotic distribution of extreme values.

This result is known as the "Fisher-Tippet theorem"; in the following I'll expose it developing its formal justification.

4.1 Fisher – Tippett theorem

Observe that $F(z)^n \to 0$ as $n \to \infty$, so that the distribution of M_n degenerate to a point mass on z_+ (the smallest value of z such that $F(z) = 0$)
This drawback is avoided by allowing a linear renormalization of the variable M_n .

$$M_n^* = \frac{M_n - b_n}{a_n}$$

If there exist a sequence of constants $a_n > 0$ and b_n such that

$$\lim_{n\to\infty} \Pr\left\{\left(M_n - b_n\right)/a_n \leq z\right\} = G\left(z\right)$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G\left(z\right) = \exp\left\{-\left[1 + \xi\left(\frac{\left(z-\mu\right)}{\sigma}\right)\right]_+^{-1/\xi}\right\}$$

$$\Pr\left\{M_n \leq z\right\} \approx G\left\{\left(z - b_n\right)/a_n\right\} = G^*\left(z\right)$$

Note that if this theorem enables approximation of $M^*{}_n$ by a member of GEV family for large n the distribution of $M^*{}_n$ itself can also be approximated by a different member of the same family. Since the parameters of the distribution have to be estimated anyway, it is irrelevant that the parameters of the distribution G are different from those of G^*
It's trivial to check that $G\left(z\right)$ summarizes, in parametric form, three distribution functions:

Frechet

$$\Phi_\alpha\left(x\right) = \begin{cases} 0, & x \leq 0 \\ \exp\left(-x^{-\alpha}\right), & x > 0 \end{cases} \qquad \alpha > 0$$

Weibull

$$\psi_\alpha\left(x\right) = \begin{cases} \exp\left(-\left(-x^{-\alpha}\right)\right), & x \leq 0 \\ 1, & x > 0 \end{cases} \qquad \alpha > 0$$

Gumbel

$$\Lambda\left(x\right) = \exp\left(-e^{-x}\right) \qquad x \in \mathbb{R}$$

The early versions of the theorem on the asymptotic distribution of maxima work with these three distinct functions and many contemporary authors conserve this approach [59].
To proof the main statement of this chapter would be interesting, to fully appreciate it's semantic, to present the argumentation originally used by Fisher-Tippett and subsequently integrated by Gnedenko; nevertheless I preferred replicate the structure proposed by De Haan [22] because it well represents the standard setup assumed by the modern EVT.

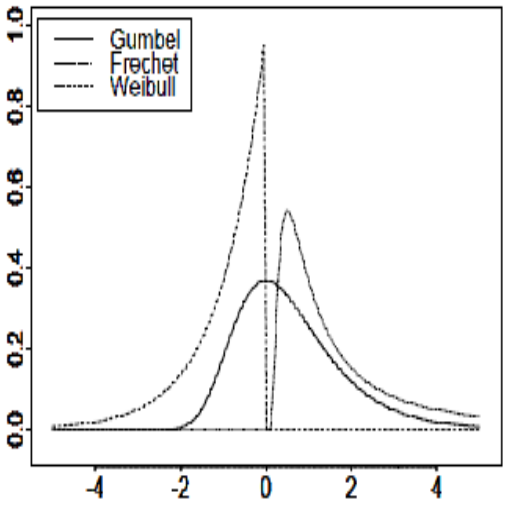

Figure 4.1: *Densities of the standard extreme value distributions with* $\alpha = 1$. - *[27], p. 122*

Preliminary to the proof are some reformulations and lemmas introduced here to make formulas easier to treat.
Let's begin observing that:

$$P\left(\max\left(X_1, ..., X_n\right) \leq x\right) = P\left(X_1 \leq x, ..., X_n \leq x\right) = F^n\left(x\right)$$

converges to zero for $x < x^*$and to 1 for $x > x^*$
Suppose there exists a sequence of constants $a_n > 0$ and b_n $(n = 1, 2, ...)$ real such that $\frac{\max(X_1,...,X_n)-b_n}{a_n}$ has a nondegenerate limit distribution as $n \to \infty$:

$$\lim_{n\to\infty} F^n\left(a_n x + b_n\right) = G\left(x\right)$$

In what follows the class of distributions F satisfying last condition is derived.

Proceed with trivial transformations:

$$\lim_{n\to\infty} n \log F\left(a_n x + b_n\right) = \log G\left(x\right)$$

$$\lim_{n\to\infty} \frac{-\log F\left(a_n x + b_n\right)}{1 - F\left(a_n x + b_n\right)} = 1$$

$$\lim_{n\to\infty} n\left(1 - F\left(a_n x + b_n\right)\right) = -\log G\left(x\right)$$

$$\lim_{n\to\infty} \frac{1}{n\left(1 - F\left(a_n x + b_n\right)\right)} = \frac{1}{-\log G\left(x\right)}$$

Working with inverse functions makes next steps simpler, so define:

$$\psi^{-1}(y) = \inf\{x : \psi(x) \geq y\}$$

Lemma 4.1.1 *Suppose f_n is a sequence of nondecreasing functions and g is a non decreasing function. Suppose that for each x in some open interval (a, b) - a continuity point of g - $\lim_{n\to\infty} f_n(x) = g(x)$.*
Let $f_n^{\leftarrow}$, $g^{\leftarrow}$ be the left-continuous inverse of f_n and g. Then, for each x in the interval $(g(a), g(b))$ - a continuity point of $g^{\leftarrow}$ - we have

$$\lim_{n\to\infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x)$$

Proof
Let x be a continuity point of $g^{\leftarrow}$ and consider an arbitrary $\varepsilon > 0$. The goal is to prove that for $n, n_0 \in \mathbb{N}$, $n \geq n_0$: $f_n^{\leftarrow}(x) - \varepsilon \leq g^{\leftarrow}(x) \leq f_n^{\leftarrow}(x) + \varepsilon$
For the left-side inequality (the right end is analogous), choose $0 < \varepsilon_1 < \varepsilon$ such that $g^{\leftarrow}(x) - \varepsilon_1$ is a continuity point of g. This is possible since the continuity points of g form a dense set. Because $g^{\leftarrow}$is continuous in x, $g^{\leftarrow}(x)$ is a point of increase for g; hence $g(g^{\leftarrow}(x) - \varepsilon_1) < x$. Choose $\delta < x - g(g^{\leftarrow}(x) - \varepsilon_1)$. Being $g^{\leftarrow}(x) - \varepsilon_1$ a continuity point of g, there exists n_0 such that $f_n(g^{\leftarrow}(x) - \varepsilon_1) < g(g^{\leftarrow}(x) - \varepsilon_1) + \delta < x$ for $n \geq n_0$. Hence $f_n^{\leftarrow}(x) - \varepsilon \leq g^{\leftarrow}(x)$ for the definition of $f_n^{\leftarrow}$.
Let the function U be the left-continuous inverse of $\frac{1}{(1-F)}$.
Recall that $\lim_{n\to\infty} n(1 - F(a_n x + b_n)) = -\log G(x)$; for the previous lemma this clause is equivalent to

$$D(x) = \lim_{n\to\infty} \frac{U(nx) - b_n}{a_n} = G^{\leftarrow}\left(e^{-1/x}\right) \qquad \forall\, x > 0) \tag{4.1}$$

Now we have necessary tools to derive the core issue:

Theorem 4.1.2 *The class of extreme value distributions is $G_\gamma(ax + b)$ with $a > 0$, b real, where $G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-\frac{1}{\gamma}}\right)$, $1 + \gamma x$ with γ real and where for $\gamma = 0$ the right-end side is interpreted as $\exp(-e^{-x})$*

Proof
Consider the class of limit functions D, suppose that 1 is a continuity point of D and note that for continuity points $x > 0$

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = D(x) - D(1) = E(x)$$

For $y > 0$ it is:

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}$$

$\lim_{t\to\infty}\frac{U(ty)-U(t)}{a(t)}$ and $\lim_{t\to\infty}\frac{a(ty)}{a(t)}$ exist. Suppose not: then there are A_1, A_2 B_1 B_2 with $A_1 \neq A_2$ or $B_1 \neq B_2$ where B_i are limit points of $\frac{U(ty)-U(t)}{a(t)}$and A_i are limit points of $\frac{a(ty)}{a(t)}$, $i = 1, 2$ as $t \to \infty$
We have: $E(xy) = E(x) A_i + B_i$
because:

$$\begin{aligned} D(xy) - D(1) &= \lim_{t\to\infty}\frac{U(txy)-b(t)}{a(t)} - \lim_{t\to\infty}\frac{U(t)-b(t)}{a(t)} \\ &= \lim_{t\to\infty}\frac{U(txy)-U(t)}{a(t)} \\ &= E(xy) \end{aligned}$$

For an arbitrary x take a sequence of continuity points x_n with $x_n \uparrow x$ $n \to \infty$; $E(x_n y) \to E(xy)$ and $E(x_n) \to E(x)$ since E is left continuous; hence $E(xy) = E(x) A_i + B_i \,\forall\, x:\ x > 0,\ \forall y:\ y > 0$

$$\begin{aligned} &E(x) A_1 + B_1 = E(x) A_2 + B_2 \\ &E(x)(A_1 - A_2) = B_2 - B_1 \end{aligned}$$

E cannot be constant (G is nondegenerate) so we have $(A_1 = A_2)$ and $(B_1 = B_2)$: $A(y) = \lim_{t\to\infty}\frac{a(ty)}{a(t)}$ exists and $E(xy) = E(x) A(y) + E(y)$ being

$$\lim_{t\to\infty}\frac{U(ty)-U(t)}{a(t)} = D(y) - D(1) = E(y)$$

Writing $s = \log x$ $t = \log y$ and $H(x) = E(e^x)$
it is: $H(t+s) = H(s) A(e^t) + H(t)$ (4.2)
or, equivalently (since $H(0) = 0$) : $\frac{H(t+s)-H(t)}{s} = \frac{H(s)-H(0)}{s} A(e^t)$

$$\lim_{s\to 0}\frac{H(t+s)-H(t)}{s} = \lim_{s\to 0}\frac{H(s)-H(0)}{s} A\left(e^t\right)$$

From the definition of derivative: $H'(t) = H'(0) A(e^t)$ (4.3)
Define: $Q(t) = H(t) \Big/ H'(0)$ (4.4)
(note that: $Q(0) = H(0) \Big/ H'(0) = 0$ and $Q'(0) = \left(H(0) \Big/ H'(0)\right)' = (H(0)/k)' = H'(0) \Big/ k = 1$)

$$\begin{aligned} Q(t+s) - Q(t) &= \frac{H(t+s)-H(t)}{H'(0)} \\ &= \frac{H(s)A(e^t)+H(t)-H(t)}{H'(0)} \qquad \text{cfr. (4.2)} \\ &= Q(s)A(e^t) \end{aligned}$$

$A\left(e^{t}\right)=\frac{H'(t)}{H'(0)}=Q'\left(t\right)$ (4.5); see (4.3) and (4.4)
hence: $Q\left(t+s\right)-Q\left(t\right)=Q\left(s\right)Q^{'}\left(t\right)$
and analogously $Q\left(t+s\right)-Q\left(s\right)=Q\left(t\right)Q^{'}\left(s\right)$

$$Q\left(t+s\right)=Q\left(t\right)+Q\left(s\right)Q^{'}\left(t\right)=Q\left(s\right)+Q\left(t\right)Q^{'}\left(s\right)$$

grouping by $Q\left(\cdot\right)$ and dividing by s : $Q\left(t\right)\frac{Q^{'}(s)-1}{s}=Q\left(s\right)\frac{Q^{'}(t)-1}{s}$

Send to limit and apply once more the definition of derivative;

$$\begin{array}{l}\lim\limits_{s\to0}Q\left(t\right)\frac{Q^{'}(s)-1}{s}=\lim\limits_{s\to0}Q\left(s\right)\frac{Q^{'}(t)-1}{s}\\ \lim\limits_{s\to0}Q\left(t\right)\frac{Q^{'}(s)-Q^{'}(0)}{s}=\lim\limits_{s\to0}\frac{Q(s)-Q(0)}{s}\left(Q^{'}\left(t\right)-1\right)\\ Q\left(t\right)Q^{''}\left(0\right)=Q^{'}\left(0\right)\left(Q^{'}\left(t\right)-1\right)\\ Q\left(t\right)Q^{''}\left(0\right)=\left(Q^{'}\left(t\right)-1\right)\end{array}$$

differentiating twice:

$$Q^{'}\left(t\right)Q^{''}\left(0\right)=Q^{''}\left(t\right)$$

Define:

$$\gamma=Q^{''}\left(0\right)=\frac{Q^{''}\left(t\right)}{Q^{'}\left(t\right)}=d\left(\log\left(Q^{'}\left(t\right)\right)\right)$$

consequently: $\log\left(Q^{'}\left(t\right)\right)=\int\gamma dt=\gamma t$ and
$Q^{'}\left(t\right)=e^{\gamma t}$ or, equivalently: $Q\left(t\right)=\int_{0}^{t}e^{\gamma s}ds$

Recalling (4.5):

$$\frac{H\left(t\right)}{H^{'}\left(0\right)}=\int_{0}^{t}e^{\gamma s}ds=\frac{1}{\gamma}\int_{0}^{t}e^{\gamma s}ds=\frac{e^{\gamma t}-1}{\gamma}$$

$H\left(t\right)=H^{'}\left(0\right)\frac{e^{\gamma t}-1}{\gamma}$ (4.6)
Remembering that: $H\left(x\right)=E\left(e^{x}\right)$ and that $D\left(x\right)-D\left(1\right)=E\left(x\right)$

$$H\left(t\right)=E\left(e^{t}\right)=D\left(e^{t}\right)-D\left(e^{1}\right)$$

Take logarithm and adapt (4.4):

$$D\left(t\right)-D\left(1\right)=E\left(t\right)=H\left(\log t\right)=H'\left(0\right)\frac{e^{\gamma\log(t)}-1}{\gamma}$$

$$D\left(t\right)=D\left(1\right)+H'\left(0\right)\frac{t^{\gamma}-1}{\gamma}$$

Apply the definition of generalized inverse function:

$$D^{\leftarrow}\left(x\right)=\left(1+\gamma\frac{x-D\left(1\right)}{H^{'}\left(0\right)}\right)$$

and from (4.1):

$$D(x) = G^{\leftarrow}\left(e^{-1/x}\right); \qquad D^{\leftarrow}(x) = \frac{1}{-\log(G(x))}$$

merging the latter two it's finally:

$$G_{\gamma}(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$$

Chapter 5

A stochastic process for extreme interest rates: Vasicek revised

As previously remarked, the aim of my PhD work is the description of a "mean reverting" process that represents the behavior of interest rates starting from non-standard assumptions about the stationary distribution of the process; in particular, the main result presented here introduces a modification of the Vasicek SDE in which the random variable follows an extreme type distribution and is consequently modeled according to the Extreme Value Theory approach. The importance of the analysis of financial variables from the EVT point of view is both theoretic - as important authors showed in recent years – and practical: during the financial crisis that we are facing, shocks in market movements systematically violate the distributional conditions underpinning classical quantitative finance.[1]
Contributions to a so actual topic come from at least two research areas each of one develops a particular aspect of the approach that I'm going to propose here; such areas can be – without demand of exhaustiveness - summarized as follows:

1. The Extreme Values applications to interest rates modeling but without consideration of stochastic processes

[1] Extreme value theory is not the first and not the only resource used to model rare events; a well known tool is represented by jump diffusion models [73] in which a Poisson process is added to the Ito diffusive process in order to better capture improvise volatility shocks in real-life rates fluctuations. In jump diffusion models the log-returns run according to the SDE: $r(t) = \mu t + \sigma W(t) + \sum_{i=1}^{M(t)} \ln(Y_i + 1)$ where $M(t)\ t \geq 0$ is a Poisson process on (Ω, F, P) with parameter $\lambda = 0$ - For an overview about the Poisson process, the reader is invited to refer to [15] The theoretical connection between jump diffusion models and EVT models is examined in [58]; the author proofs that if $J = \sum_{i=1}^{M(t)} \ln(Y_i + 1)$ is in the Maximum Domain of Attraction of $\Phi_{\alpha\pm}$, a random variable of the type $N(\mu, \sigma^2\tau) + J$ is in the Maximum Domain of Attraction of $\Phi_{\alpha\pm}$ again.

2. the Extreme Values theoretical field extended to stochastic processes but without explicit applications

The main result presented here can be intended as a previously unexplored link between these neighbor areas; in next paragraphs I'll introduce a general view of both of them to better understand how my bridging idea is inserted and connected to its supporting sides

5.0.1 Financial applications of Extreme Value Theory

In the last ten years the amount of papers describing financial applications of Extreme Value Theory is substantially increased under the pressure of recognized anomalies in financial returns distribution. Getting a look to the paper's bibliography it's easy to check the recurrence of various monographs (for example [26], [27], [28], [56]) that we can assume as "templates" for the subsequent researches.
All of them have as central point the Value at Risk, a standard in risk evaluation that measures risk in terms of percentile of a loss distribution.
VaR concerns "extreme" percentiles, so that it seems a natural choice to work with those distributions, as EVT distributions, that better capture the behavior of financial variables at the tail of the mass function.
In the study of heavy negative tailed distributions which are encountered during empirical examination of asset log returns, EVT is a basis for the so-called Peak Over Threshold Method (POT). The POT method examines the distribution of excesses over a sufficiently high limit:

$$F_u(y) = P\{X - u = y | X > u\}$$

A "sufficiently high threshold" is a threshold that provides an optimal balance between the bias of the model which is increased as the threshold becomes lower and the variance of it which grows as the threshold does due to the lack of data points; details on its calculation will be discussed later.
According to Pickands, Balkema and De Haan the excess distribution can be approximated well by a general Pareto distribution as the threshold becomes large:

$$\lim_{z\to\infty} F_z(x) = G_{\xi,\tilde{\alpha}(z)}(x)$$

with:

$$G_{\xi,\tilde{\alpha}(z)}(x) = \left[1 + \frac{\xi y}{\tilde{\alpha}}\right]^{-1/\xi} \text{ and } \tilde{\sigma} = \sigma + \xi(z - \mu)$$

A sketch of the proof of this fact is supplied in [14]:

$$G^n(z) \approx \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\alpha}\right)\right]^{-1/\xi}\right\}$$
$$n \log G(z) \approx -\left[1 + \xi\left(\frac{z-\mu}{\alpha}\right)\right]^{-1/\xi}$$

a Taylor series expansion entails:

$$\log G(z) \approx -\{1 - F(z)\}$$
$$1 - G(z) \approx \frac{1}{n}\left[1 + \xi\left(\frac{z-\mu}{\alpha}\right)\right]^{-1/\xi}$$
$$1 - G(z+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{z+y-\mu}{\alpha}\right)\right]^{-1/\xi}$$
$$\Pr\{X > z + y | X > z\} \approx \frac{n^{-1}[1+\xi(z+y-\mu)/\alpha]^{-1/\xi}}{n^{-1}[1+\xi(z-\mu)/\alpha]^{-1/\xi}} = \left[\frac{1+\xi(z-\mu)/\alpha+\xi(y)/\alpha}{1+\xi(z-\mu)/\alpha}\right]^{-1/\xi} =$$
$$= \left[1 + \frac{\xi y}{\alpha+\xi(z-\mu)}\right]^{-1/\xi}$$

$$= \left[1 + \frac{\xi y}{\tilde{\alpha}}\right]^{-1/\xi} \text{for } \tilde{\sigma} = \sigma + \xi(z-\mu)$$

The GPD parameter estimation is achieved via momentum method as in [41]; estimators are:

$$\xi = \frac{1}{2}\left[\frac{\mu^2}{\sigma^2} - 1\right] \text{ and } \tilde{\alpha} = \frac{1}{2}\mu\left[\frac{\mu^2}{\sigma^2} + 1\right]$$

As previously announced the use of GPD functional form is legitimated only for "maxima" i.e. only for values over a threshold, but maxima are defined only in terms of such a functional form; consequently the threshold selection is made critical from this self-referential loop.
To exit the loop a criterion of eligibility is given by the "stationarity" of the model parameters over a threshold: if a GPD is a reasonable model for excesses over a threshold z_0, then excesses of a higher threshold z should also follow a GPD; the shape parameters of the two distributions are identical.
Denoting by σ_z the value of generalized Pareto scale parameter for a threshold $z > z_0$, it is:

$$\sigma_z = \sigma_{z_0} + \xi(z - z_0)$$

so that the scale parameter changes with z unless $\xi = 0$; this shortcoming can be remedied by reparametrizing the Generalized Pareto scale parameter as

$$\sigma^* = \sigma_z - \xi z$$

which is constant with respect to z. Consequently, estimates of both σ^* and ξ should be constant above z_0, if z_0 is a valid threshold for excesses. Sampling variability means that the estimates of these quantities will not be exactly constant, but they should be stable after allowance for their sampling errors. This argument suggest plotting both $\hat{\sigma}^*$ and $\hat{\xi}^*$ against z_0 as the lowest value of z for which the estimates remain near-constant.
Once completed the parameter estimation and fitted the sample, the VaR is the tail quantile for a given probability p, available inverting the GPD function:

$$\hat{q}_p = \frac{\tilde{\alpha}}{\hat{\xi}}\left(p^{-\hat{\xi}} - 1\right)$$

This GPD analytic framework is adopted for example by [36] to produce a model of the extreme movements of Turkish interest rates during the 2001 financial

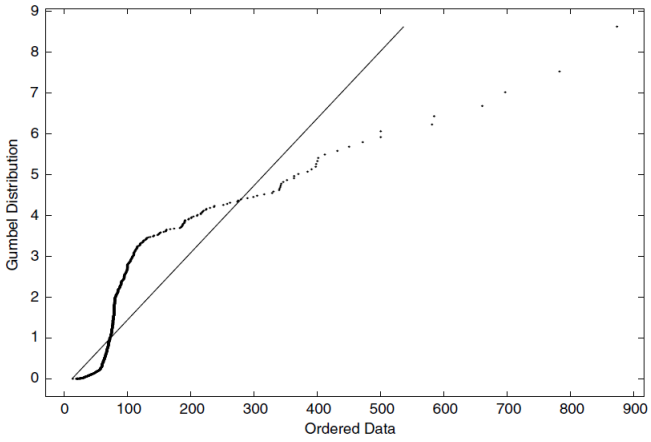

Figure 5.1: A QQ-plot (quantile–quantile plot) is a visual tool to examine whether a sample comes from a specific distribution. Specifically, the quantiles of a hypothesized distribution are plotted against the quantiles of an empirical distribution. If the sample comes from the hypothesized distribution, the QQ-plot is linear. - [36] pag. 555

crisis; the results were remarkable as noted from the authors: "Our estimation results from the pre-crisis data indicate that every 4 years one could expect to see a day with overnight interest rates over 1000 percent (simple annual). In other words, the extraordinary levels observed during the crisis were in the nature of the economy before they actually materialized."

5.0.2 Extreme Value in stochastic processes

The probably most complete and referenced source about this topic is in [11] and related works; in those studies the stochastic processes of extreme values are examined in a risk management context but, as anticipated, an applicative framework is not developed.

These works contain a composite set of very interesting results; I give here a synopsis of the most relevant ones from my current point of view.

As usual we have the diffusion:

$$dX_t = \mu\left(X_t\right) dt + \sigma\left(X_t\right) dW_t$$

Associated with it are the *scale* function s and the *speed* measure m.

The first one is defined as:

$$s\left(x\right) = \int_z^x \exp\left\{-2\int_z^y \frac{\mu\left(x\right)}{\sigma^2\left(x\right)} dt\right\} dy$$

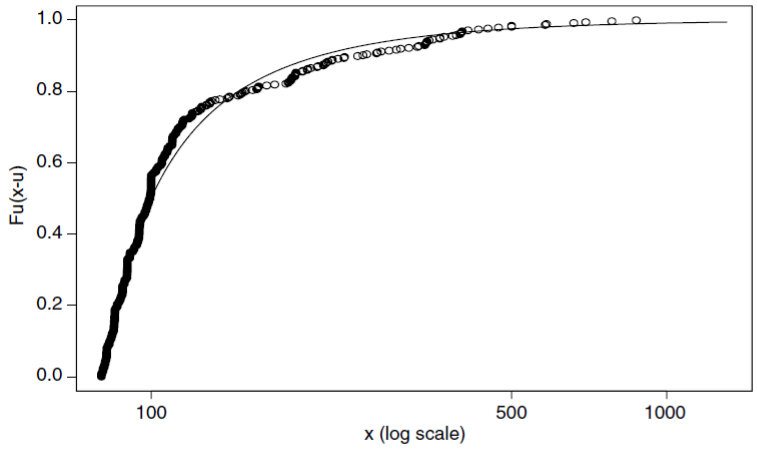

Figure 5.2: Fitting empirical distribution (circles) with GPD (lines) - [36] pag. 556

Figure 5.3: Threshold individuation - [36] pag. 556

the speed measure is absolutely continuous with Lebesgue density:

$$m^{'}(x) = \frac{2}{\sigma^2(x)\, s^{'}(x)}$$

In this situation (x_t) is ergodic and its stationary distribution is absolutely continuous with Lebesgue density

$$h(x) = m^{'}(x) \Big/ |m|$$

For any initial value$(X_0) = y \in (l, r)$ and any $u_t \uparrow r$

$$\lim_{t\to\infty} \left| P^y\left(M_t^X \leq u_t\right) - F^t(u_t)\right| = 0$$

where $F$is a df , defined by

$$F(x) = e^{-1/(|m|s(x))} I_{(z,r)}(x)\, \forall z \in (l, r)$$

from Taylor expansion and the fact that $\lim_{x\uparrow r} s(x) = \infty$

$$\bar{F}(x) \sim \left(|m| \int_z^x s^{'}(y)\, dy\right)^{-1} \sim (|m|\, s(x))^{-1}$$

Lemma 5.0.3 *Let μ and σ^2 be differentiable function on (x_0, r) for some $x_0 < r$ such that:*

$\lim_{x\uparrow r} \frac{d}{dx}\left\{\frac{\sigma^2(x)}{\mu(x)}\right\} = 0$ *and* $\lim_{x\uparrow r} \frac{\sigma^2(x)}{\mu(x)} \exp\left\{-2\int_z^x \frac{\mu(t)}{\sigma^2(t)} dt\right\} = -\infty$

Then

$$\bar{F}(x) \sim |\mu(\mathrm{x})|\, h(x)$$

where $h(x)$ is the stationary density of X_t

Proof:

$$s^{''}(x) = -2s^{'}(x)\, \frac{\mu(x)}{\sigma^2(x)} x \in (l, r)$$

$\lim_{x\uparrow r} \frac{2\int_z^x s^{'}(y) dy}{-s^{'}(x)\mu(x)/\sigma^2(x)} = \lim_{x\uparrow r} \frac{2s^{'}(y)}{-s^{'}(x)o(1) - s^{''}(x)\mu(x)/\sigma^2(x)} = 1$ from De l'Hospital

$$\bar{F}(x) \sim 2\mu(\mathrm{x}) / |\mathrm{m}| s^{'}(x)\, \sigma^2(x)$$

that compared with scale measure and speed function gives:

$$\bar{F}(x) \sim |\mu(\mathrm{x})|\, h(x)$$

If we apply this relation to the solution of Vasicek SDE

$$X_t = \frac{c}{d} + \left(x - \frac{c}{d}\right) e^{-dt} + \sigma \int_0^t e^{-d(t-s)} dW_s$$

knowing that $X_{\mathrm{t}} \sim \mathrm{N}\left(\frac{\mathrm{c}}{\mathrm{d}}, \frac{\sigma^2}{2d}\right)$

we have $\bar{F}(x) \sim \mathrm{d}\frac{(\mathrm{x}-\mathrm{c}/\mathrm{d})^2}{\sigma^2/2d} \bar{H}(x)$

where $\bar{H}(x)$ is the tail of stationary normal distribution function.

5.1 Extreme Value Vasicek process

Now it is time to collect together the items described in previous chapters to write down the process with needed charactheristics; observe that the mathematical device proposed hereafter can be easily generalized so to build processes that are functions of arbitrarily chosen stationary distributions.

Theorem 5.1.1 .

Let

$$dx_t = -\gamma (x_t - \alpha) dt + \sigma (x_t) dW_t$$

be the Vasicek model representing the time evolution of the instantaneous interest rate (with the usual conventions on the meaning of the terms).

If $f(x,t) = \exp(-e^{-x} - x)$ *is the conditional density of the stochastic process* dx_t *, then*

$$\sigma^2(x) = 2\left[(c - kx)\ e^x - \exp\ \left(e^{-x} + x\right)\left(\ k\Gamma\left(0, e^{-x}\right) - \frac{h}{2}\right)\right]$$

holds with $\Gamma(0,u) = \int u^{-1} e^{-u} du$

Proof:

A fundamental PDE from the physics of particles, relates the conditional density of the stochastic process $dx_t = \mu(x_t, t)\,dt + \sigma(x_t, t)\,dW_t$ to the drift and the variance terms of the same process [33] [79].
This PDE is the Fokker-Planck equation; it states that:

$$\frac{1}{2}\frac{\partial^2\left(\sigma^2 f\right)}{\partial x^2} - \frac{\partial(\mu f)}{\partial x} = \frac{\partial f}{\partial t}$$

Because the relevance, for this work, of the *stationary* distribution of the process, $f(x,t)$ is constant so that it can be derived from $\frac{1}{2}\frac{\partial^2\left(\sigma^2 f\right)}{\partial x^2} - \frac{\partial(\mu f)}{\partial x} = 0$ through the following considerations:
Integrating the previous equation it becomes:

$$\frac{\partial}{\partial x}\left(\frac{\sigma^2(x)}{2} f\right) - \mu(x) f = \frac{1}{2} k_1$$

it can be rewritten as:

$$\frac{\partial}{\partial x}\left(\frac{\sigma^2(x)}{2} f\right) - \frac{2\mu(x)}{\sigma^2(x)}\frac{\sigma^2(x) f}{2} = \frac{1}{2} k_1$$

so that it is clear that

$$s(x) = \exp\left(-\int \frac{2\mu(v)}{\sigma^2(v)} dv\right) \tag{5.1}$$

can be assumed as *integrating factor* giving:

$$\frac{\partial}{\partial x}\left(s(x)\,\sigma^2(x) f\right) = s(x)\, c_1$$

a further integration returns:

$$f = \frac{c_1 \int s(v)\,dv + c_2}{s(x)\,\sigma^2(x)}$$

c_1 and c_2 must be determined so that f – a density – be not negative for whatever value of x; a convenient choice is: $c_1 = 0$ and $c_2 \geq 0$ obtaining:

$$f = \frac{1}{|m|\, s(x)\,\sigma^2(x)}$$

$$s(x) = \frac{1}{|m|\,\sigma^2(x) f} \tag{5.2}$$

But from (5.1) it is known that:

$$\frac{\partial}{\partial x}\ln(s(x)) = -\frac{2\mu(v)}{\sigma^2(v)}$$

and hence that:

$$\mu(x) = -\frac{\sigma^2(x)\,\frac{\partial}{\partial x}\ln(s(x))}{2}$$

substituting $s(x)$ with (5.2)

$$\mu(x) = -\frac{\sigma^2(x)\,\frac{\partial}{\partial x}\ln\left(\frac{1}{|m|\sigma^2(x)f}\right)}{2}$$

$$= -\frac{1}{2}\sigma^2(x)\left(-\frac{\partial}{\partial x}\ln(f(x)) - \frac{\partial}{\partial x}\ln\left(\sigma^2(x)\right) - \frac{\partial}{\partial x}\ln(|m|)\right) =$$
$$= -\frac{1}{2}\sigma^2(x)\left(-\frac{\partial}{\partial x}\ln(f(x)) - \frac{\partial}{\partial x}\ln\left(\sigma^2(x)\right)\right) =$$
$$= \frac{1}{2}\sigma^2(x)\left(\frac{\partial}{\partial x}\ln(f(x)) + \frac{\frac{\partial}{\partial x}\left(\sigma^2(x)\right)}{\sigma^2(x)}\right) \tag{5.3}$$

The last expression is a linear first order non-homogeneous differential equation whose solution is:

$$\sigma^2(x) = 2e^{-\int \frac{\partial}{\partial x}\ln(f(x))dx}\left(\int \mu(x)\, e^{\int \frac{\partial}{\partial x}\ln(f(x))dx}dx + h\right) \tag{5.4}$$

$$\sigma^2(x) = 2e^{-\ln(f(x))}\left(\int \mu(x)\, e^{\ln(f(x))}dx + h\right) \tag{5.5}$$

$$\sigma^2(x) = \frac{2}{f(x)}\left(\int \mu(x)\, f(x)\, dx + h\right) \tag{5.6}$$

Focusing the attention on a specific *mean reverting* interest rate model, properly the Vasicek model

$$dX_t = (c - kX_t)\, dt + \sigma(X_t)\, dW$$

define:

$$\mu(x) = (c - kX_t)$$

Furthermore, because a normally distributed random variable belongs to the domain of attraction of the Gumbel distribution and from the fact that in the Vasicek process the random variable has normal stationary distribution, it is verified that

$$f = \exp\left(-e^{-x} - x\right)$$

Hence, recalling (5.3)

$$\sigma^2(x) = \frac{2}{\exp\left(-e^{-x} - x\right)}\left(\int (c - kx)\exp\left(-e^{-x} - x\right)dx + h\right)$$

$$\sigma^2(x) = 2\exp\left(e^{-x} + x\right)\left(c\exp\left(-e^{-x}\right) - k\int x\exp\left(-e^{-x} - x\right)dx + h\right) \tag{5.7}$$

To solve the last integral in the formula (5.6) proceed integrating by parts after an opportune substitution:

$u = e^{-x}$
$x = -\ln(u)$
$dx = -1/u$

$\int x \exp(-e^{-x} - x)\, dx = \int \ln(u)\, e^{-u} du = -e^{-u} \ln(u) + \int u^{-1} e^{-u} du = \Gamma(0, u) - e^{-u} \ln(u) =$
$= \Gamma(0, e^{-x}) + x \exp(-e^{-x})$

where $\Gamma(0, u) = \int u^{-1} e^{-u} du$

is the Gamma incomplete function.

After a substitution in (5.7)

$$\sigma^2(x) = 2 \exp\left(e^{-x} + x\right) \left(c \exp\left(-e^{-x}\right) - k\left(\Gamma\left(0, e^{-x}\right) + x \exp\left(-e^{-x}\right)\right) + h\right)$$

that simplified is:

$$\sigma^2(x) = 2\left[(c - kx)\ e^x - \exp\ \left(e^{-x} + x\right)\left(\ k\Gamma\left(0, e^{-x}\right) - \frac{h}{2}\right)\right]$$

Appendix A.5 – Derivation of the Fokker-Planck equation

Suppose we have an arbitrary smooth function f. Recalling that Ito integrals have mean 0, it follows from Ito's lemma that

$$E\left[f\left(X_t\right)\right] = E\left[f\left(X_0\right) + \int_0^t \left(f^{'}\left(X_u\right) a\left(X_u\right) + \frac{1}{2} f^{''}\left(X_u\right) \sigma^2\left(X_u\right)\right) du\right]$$

Differentiating with respect to time, we find that

$$\frac{d}{dt} E\left[f\left(X_t\right)\right] = E\left[f^{'}\left(X_u\right) a\left(X_u\right) + \frac{1}{2} f^{''}\left(X_u\right) \sigma^2\left(X_u\right)\right]$$

Now, setting $X_0 = x_0$, the definition of expected value gives:

$$\int f\left(x\right) \frac{\partial}{\partial t} p\left(x,\, t|\, x_0, 0\right) dx = \int \left(f^{'}\left(X_u\right) a\left(X_u\right) + \frac{1}{2} f^{''}\left(X_u\right) \sigma^2\left(X_u\right)\right) p\left(x,\, t|\, x_0, 0\right) dx$$

where the integral is over all nonzero values of $p\left(x,\, t|\, x_0, 0\right)$.

We can hence integrate the right-hand side by parts:

$$\begin{array}{l} \int f\left(x\right) \frac{\partial}{\partial t} p\left(x,\, t|\, x_0, 0\right) dx \\ = \int f\left(x\right) \left(-\frac{\partial}{\partial x}\left(a\left(x\right) p\left(x,\, t|\, x_0, 0\right)\right) + \frac{1}{2} \frac{\partial^2}{\partial x^2}\left(\sigma^2\left(x\right) p\left(x,\, t|\, x_0, 0\right)\right)\right) dx \end{array}$$

and conclude that the transition density satisfies:

$$\frac{\partial}{\partial t} p\left(x,\, t|\, x_0, 0\right) = -\frac{\partial}{\partial x}\left(a\left(x\right) p\left(x,\, t|\, x_0, 0\right)\right) + \frac{1}{2} \frac{\partial^2}{\partial x^2}\left(\sigma^2\left(x\right) p\left(x,\, t|\, x_0, 0\right)\right)$$

Quantitative finance offers a famous example of explicit derivation of the volatility term from density through the Fokker-Planck equation in Dupire's analysis of volatility smiles [24].
Dupire overcomes the limits of constant volatilities in classical Black-Scholes-Merton model making the volatility term stochastic and sensitive to the price level; he compensates the distorsive effects of costant model volatility extracting a volatility convex function from Fokker-Planck equation, that is:

$$\sigma\left(K, T\right) = \left(\frac{2 \frac{\partial C}{\partial t}\left(K, T\right)}{K^2 \frac{\partial^2 C}{\partial K^2}\left(K, T\right)}\right)^{\frac{1}{2}}$$

where:

T: expiration time of a call option C written on a not-specified underlying security

K: strike price of the call option C

$\sigma\left(K, T\right)$: volatility of the call option C

Chapter 6

Implementation issues

6.1 Euler scheme application

To test the quality of the EVT interest rates model against experience, the experimental section of this research requires the simulation of values from the same model.
To explain the simulation approach with an adequate level of generality, take the usual process $dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t$ or, better, take a *discretized* version $\left\{\hat{X}_h, \hat{X}_{2h}, ..., \hat{X}_{mh},\right\}$ of it where m is the number of time steps, h is a constant and $mh = T$. The smaller the value of h, the closer our discretized path will be to the continuous-time path we wish to simulate.
Of course this will be at the expense of greater computational effort.
While there are a number of *discretization schemes* available, I will illustrate the simplest and perhaps most common scheme, the *Euler* scheme.
The Euler scheme is intuitive, easy to implement and satisfies

$$\hat{X}_t = \hat{X}_{(t-1)} + \mu\left(t, \hat{X}_{(t-1)}\right) h + \sigma\left(t, \hat{X}_{(t-1)}\right) \sqrt{h} Z$$

where the Z's are IID $N(0;\, 1)$.

To estimate X_t using the Euler scheme, then for a fixed number of paths, n, and discretization interval, h, we have the following algorithm:

for $j = 1$ **to** n
    set $t = 0; \hat{X} = X_0$
    for $k = 1$ **to** $T/h = m$
        generate $Z \sim \mathrm{N}(0, 1)$

set $\hat{X}_t = \hat{X}_{(t-1)} + \mu\left(t, \hat{X}_{(t-1)}\right) h + \sigma\left(t, \hat{X}_{(t-1)}\right) \sqrt{h} Z$
set $t = t + h$
end for
set $f_j = f(\hat{x})$
end for
set $\hat{\theta}_n = (f_1 + ... + f_n)/n$
set $\hat{\sigma}_n^2 = \sum_{j=1}^{n} \left(f_j - \hat{\theta}_n\right)^2 / (n-1)$
set ***Confidence_Limit*** $= \hat{\theta}_n \pm z_\alpha \frac{\hat{\sigma}_n}{\sqrt{n}}$

Clearly the procedure is affected by a discretization error related to the size of discretization step.
The discretization error may be defined by $D = \left|X_T - \hat{X}_T\right|$ and it is very important when simulating SDE's to ensure that D is suffciently small.
A common method of controlling discretization error is as follows:
Let be $\hat{X}_T^m$ our estimator of X_T when we use m discretization points. We first compute $\hat{X}^m$ and $\hat{X}^{2m}$ for a reasonably large value of m. If $\left|\hat{X}^m - \hat{X}^{2m}\right|$ is sufficiently small then we can assume that $2m$ is a sufficiently large sample size to guarantee a negligible discretization error.

6.1.1 Improving the Euler scheme: the Milstein scheme

To improve the precision of the Euler scheme we could include higher-order terms in a Taylor-like expansion of $\mu(X(t))$ paying attention to its compatibility with the rules of Ito calculus rather than ordinary calculus.
Rewrite the Ito process in integral form:

$$X(t) = X(0) + \int_0^t \mu(X(u))\, du + \int_0^t \sigma(X(u))\, dW(u)$$

The Euler scheme results from the approximations

$$\int_t^{t+h} \mu(X(u))\, du \approx \mu(X(t))\, h$$

and

$$\int_t^{t+h} \sigma(X(u))\, dW(u) = \sigma(X(t))\, [W(t+h) - W(t)] \tag{6.1}$$

that approximate the integrands over $[t, t+h]$ by their value at t.
Now we improve the approximation of $\sigma(X(u))$ improving its accuracy over an interval $[t, t+h]$.
Ito's formula gives:

$$d\sigma(X(u)) = \left[\sigma'(X(t))\, \mu X(t) + \tfrac{1}{2}\sigma''(X(t))\, \sigma^2(X(t))\right] dt + \sigma'(X(t))\, \sigma X(t)\, dW(t)$$

Applying the Euler approximation to the process $\sigma(X(t))$ $t \leq u \leq t+h$

$$\begin{aligned}
&\sigma(X(u)) \approx \sigma(X(t)) + d\sigma(X(u)) \\
&= \sigma(X(t)) + \left[\sigma'(X(t))\mu X(t) + \tfrac{1}{2}\sigma''(X(t))\sigma^2(X(t))\right][u-t] \\
&+\sigma'(X(t))\sigma X(t)[W(u) - W(t)]
\end{aligned}$$

Dropping the higher-order term yields the approximation $\sigma(X(u)) \approx \sigma'(X(t))\sigma X(t)[W(u) - W(t)]$. Having this approximation at our disposal, we can replace (5.7) with:

$$\begin{aligned}
&\int_t^{t+h} \sigma(X(u))\, dW(u) \\
&\approx \int_t^{t+h} \left(\sigma(X(t))\sigma'(X(t))\sigma(X(t))[W(u) - W(t)]\right) dW(u) \\
&= \sigma(X(t))[W(t+h) - W(t)] \\
&+\sigma'(X(t))\sigma(X(t))\left(\int_t^{t+h}[W(u) - W(t)]\, dW(u)\right)
\end{aligned} \tag{6.2}$$

Integral in (6.1) can be written as:

$$\begin{aligned}
&\int_t^{t+h}[W(u) - W(t)]\, dW(u) = \\
&\int_t^{t+h} W(u)\, dW(u) - W(t)\int_t^{t+h} dW(u) \\
&= Y(t+h) - Y(t) - W(t)[W(t+h) - W(t)]
\end{aligned} \tag{6.3}$$

with

$$Y(t) = \int_0^t W(t)\, dW(t)$$

and

$$dY(t) = W(t)\, dW(t)$$

The solution to this SDE is, as argued in chap. 2:

$$Y(t) = \frac{1}{2}W(t)^2 - \frac{1}{2}t$$

Making this substitution in (6.2) and simplifying, we get

$$\int_t^{t+h}[W(u) - W(t)]\, dW(u) = \frac{1}{2}[W(t+h) - W(t)]^2 - \frac{1}{2}h$$

that is, using it in (6.2):

$$\begin{aligned}
&\int_t^{t+h} \sigma(X(u))\, dW(u) \\
&\approx \sigma(X(t))[W(t+h) - W(t)] \\
&+\tfrac{1}{2}\sigma'(X(t))\sigma(X(t))\left([W(t+h) - W(t)]^2 - h\right)
\end{aligned}$$

This approximation affects X(t + h); the one-step Euler approximation is hence refined in this way:

$$\begin{aligned}
&X(t+h) \approx X(t) + \mu(X(t))h + \sigma(X(t))[W(t+h) - W(t)] \\
&+\tfrac{1}{2}\sigma'(X(t))\sigma(X(t))\left([W(t+h) - W(t)]^2 - h\right)
\end{aligned}$$

The simulation algorithm becomes:

$$\hat{X}_t = \hat{X}_{(t-1)} + \mu\left(t, \hat{X}_{(t-1)}\right) h + \sigma\left(t, \hat{X}_{(k-1)h}\right) \sqrt{h} Z$$
$$+\tfrac{1}{2}\sigma'\left(t, \hat{X}_{(t-1)}\right) \sigma\left(t, \hat{X}_{(t-1)}\right) h\left(Z^2 - 1\right)$$

This solution is due to Milstein and known as "Milstein scheme" [54].

6.1.2 Variance reduction

Variance reduction techniques are useful in reducing a big amount of computational resources in numerical simulations; they attempt to limit algorithmic redoundancies making use of statistical properties of the modeled domain.

The method of control variates is among the most effective and broadly applicable techniques for improving the efficiency of Monte Carlo simulation. [37]

It exploits information about the errors in estimates of known quantities to reduce the error in an estimate of an unknown quantity. To do so, it carries out in parallel two simulations using the same random number streams and the same timestep; the first is used to obtain an estimate f_A^* of the unknown quantity, while the second is used to obtain an estimate f_B^* of the known quantity, used, via the comparison with the theoretical quantity f_A^*, to evaluate the theoretical gap between f_A and f_B.

A better estimate f_A of the quantity A is then calculated using the formula: $f_A = f_A^* - f_B^* + f_B$ where f_B is the true known value of B deducted analytically.

Turning back to our model with stochastic interest rates, we consider bonds as the underlying assets of an interest rate model so to take bond prices as a source of control variates.

Given the short rate r_i, a bond maturing at time T has initial price

$$B(0,T) = E\left[\exp\left(-\int_0^T r(u)\, du\right)\right]$$

Because $B(0,T)$ is known, we can assume $\exp\left(-\int_0^T r(u)\, du\right)$ as control variate. Using

$$\exp\left(-\frac{1}{n}\sum_{i=1}^{n} r(t_i)\right)$$

in a control variate estimator could entail some bias related to the partition size of $[0,T]$; nevertheless when an exact joint simulation of $r(t_i)$ and its time integral $Y(t_i) = \int_0^{t_i} r(u)\, du$ is allowed, it provides a bias-free control variate because $E[\exp(-Y(t_i))] = B(0,T)$.

6.2 Model calibration

In chapter 3 I gave an illustration of the Vasicek model and the technicalities related to its parameters calibration.
The maximum-likelihood rationale underpinning that context is basically transferable to the EVT – Vasicek model examined here except the need for a numerical procedure in estimating model parameters; EVT – Vasicek model differs from classical-Vasicek in the volatility term that is not a constant but a function that makes least squares unfeasible with closed forms.
For this reason optimization tasks get critical relevance in performing calibration so that a look through the optimization tool can help in understanding and adjusting model outcomes.

6.2.1 The Levenberg-Marquardt algorithm

In fitting a function $\hat{y}(t;w)$ of an independent variable t and a vector of n parameters w to a set of data points (t_i, y_i), it is customary and convenient to minimize the sum of the weighted squares of the errors (or weighted residuals) between the measured data $y(t_i)$
and the curve-fit function $y(t_i;w)$. This scalar-valued goodness-of-fit measure is called the chi-squared error criterion.

$$\begin{aligned}\chi^2(w) &= \textstyle\sum_{i=1}^{m} \frac{1}{2}\left[\frac{y(t_i)-y(t_i;w)}{\sigma_i}\right]^2 \\ &= \tfrac{1}{2}\left(Y-\hat{Y}(w)\right)^T S\left(Y-\hat{Y}(w)\right) \\ &= \tfrac{1}{2}Y^T SY - Y^T S\hat{Y} + \tfrac{1}{2}\hat{Y}^T S\hat{Y}\end{aligned} \tag{6.4}$$

The value w_i is a measure of the error in measurement $y(t_i)$. The matrix S is diagonal with $S_{ii} = 1/\sigma_i^2$
If the function $\hat{y}$ is nonlinear in the model parameters w, then the minimization of χ^2 with respect to the parameters must be carried out iteratively. The goal of each iteration is to find a perturbation h to the parameters w that reduces χ^2.
The steepest descent method is a general minimization method which updates parameter values in the direction opposite to the gradient of the objective function. It is recognized as a highly convergent algorithm for finding the minimum of simple objective functions.
For problems with thousands of parameters, gradient descent methods may be the only viable method.
The gradient of the chi-squared objective function with respect to the parameters is:

$$\begin{aligned}\tfrac{\partial}{\partial w}\chi^2 &= \left(Y-\hat{Y}(w)\right)^T S\tfrac{\partial}{\partial p}\left(Y-\hat{Y}(w)\right) \\ &= -\left(Y-\hat{Y}(w)\right)^T S\left[\tfrac{\partial\hat{Y}(w)}{\partial w}\right] \\ &= -\left(Y-\hat{Y}(w)\right)^T SJ\end{aligned}$$

where the $n \times m$ Jacobian matrix $\frac{\partial \hat{Y}}{\partial w}$ represents the local sensitivity of the function $\hat{Y}$to variation in the parameters w. For notational simplicity J will be used for $\frac{\partial \hat{Y}}{\partial w}$.
The perturbation h that moves the parameters in the direction of steepest descent is given by

$$h = -\left(Y - \hat{Y}(w)\right)^T SJ$$

where the positive scalar α determines the length of the step in the steepest-descent direction.
The Gauss-Newton method is a method of minimizing a sum-of-squares objective
function. It presumes that the objective function is approximately quadratic in the parameters
near the optimal solution. For more moderately-sized problems the Gauss-Newton method typically converges much faster than gradient-descent methods.
The function evaluated with perturbed model parameters may be locally approximated through a first-order Taylor series expansion.

$$Y(w+h) \approx \hat{Y}(w) + \left[\frac{\partial \hat{Y}(w)}{\partial w}\right] h = \hat{Y}(w) + Jh$$

Substituting the approximation for the perturbed function, $\hat{Y}(w) + Jh$ for $\hat{Y}$ in equation (5.1),

$$\chi^2(w+h) = \frac{1}{2} Y^T SY - Y^T S\hat{Y} + \frac{1}{2}\hat{Y}^T S\hat{Y} - \left(Y - \hat{Y}\right)^T SJh + \frac{1}{2} h^T J^T SJh$$

This shows that χ^2 is approximately quadratic in the perturbation h, and that the Hessian of the chi-squared fit criterion is approximately $J^T SJ$.
The perturbation h that minimizes χ^2 is found from $\frac{\partial \chi^2}{\partial w} = 0$.

$$\frac{\partial}{\partial w}\chi^2(w+h) \approx -\left(Y - \hat{Y}(w)\right)^T SJ + h^T J^T SJ$$

and the resulting normal equation for the Gauss-Newton perturbation is:

$$\left[J^T SJ\right] h = J^T S\left(Y - \hat{Y}(w)\right)$$

The Levenberg-Marquardt algorithm adaptively varies the parameter updates between the gradient descent and Gauss-Newton update,

$$\left[J^T SJ + \lambda I\right] h = J^T S\left(Y - \hat{Y}(w)\right) \tag{6.5}$$

where small values of the algorithmic parameter λ result in a Gauss-Newton update and large values of λ result in a gradient descent update. At a large distance from the function minimum, the steepest descent method is utilized to provide steady and convergent progress toward the solution. As the solution approaches the minimum, λ is adaptively decreased, the Levenberg-Marquardt

method approaches the Gauss-Newton method, and the solution typically converges rapidly to the local minimum.

The outlined algorithm has the disadvantage that if the value of λ is large, the calculated Hessian matrix is not used at all; in such cases the second derivative deliveries an information useful to scale each component of the gradient according to the curvature. This should result in *larger* movement along the directions where the gradient is *smaller*; for this reason Marquardt replaced the identity matrix in (6.4) with the diagonal of the Hessian resulting in the Levenberg-Marquardt update rule.

$$\left[J^T S J + \lambda diag\left(J^T S J\right)\right] h = J^T S\left(Y - \hat{Y}\left(w\right)\right)$$

In practical implementation Jacobian is numerically evaluated using backwards differences,

$$J_{i,j} = \frac{\partial \hat{y}_i}{\partial w_j} = \frac{\hat{y}\left(t_i; w + \delta w_j\right) - \hat{y}\left(t_i; w\right)}{\|\delta w_j\|}$$

where the j-th element of δw_j is the only non-zero element and is set to $\varepsilon_w = (1 + |w_j|)$.

If in an iteration $\chi^2\left(w\right) - \chi^2\left(w + h\right) > \varepsilon h^T\left(\lambda h + J^T S\left(y - \hat{y}\right)\right)$ then p+h is sufficiently better than w, w is replaced by w+h, and λ is reduced by a factor of ten. Otherwise λ is increased by a factor of ten, and the algorithm proceeds to the next iteration. Convergence is achieved if $\max\left(|h_i/p_i|\right) < \varepsilon_h$, $\chi^2/m < \varepsilon_y$, or $\max\left(\left|J^T W\left(Y - \hat{Y}\right)\right|\right) < \varepsilon_z$. Otherwise, iterations terminate when the iteration number exceeds a pre-specified limit.

6.2.2 Bayesian parameters inference

Least square approach to inference is well known to be affected by overfitting: the training sample could be not representative of the full population's behavior and too detailed models have the shortcoming to tradeoff a low in-sample approximation error with high variance – i.e. not acceptable out-of-sample (prediction) error.

Maximum likelihood criterion is reliable only when observed samples are the most likely; it is the most probable circumstance, but it is not certain. Conversely, a way to take into account parameters' uncertainty is to consider the maximum probability of parameters given data, parameters themselves following a distribution law.

Bayesian learning is based on this idea deriving posterior parameters distribution from the Bayes rule; given:

$P\left(y|\, w\right)$= the likelihood of the model output $y$given the set of model parameters w

$P\left(w\right)$= the prior probability of the model parameters w

the probability of parameters given the observation (the "right" way to think

about parameter estimation) is:

$$P(w|y) = \frac{P(y|w)\,P(w)}{\int P(y|w)\,P(w)}$$

In models calibration bayesian approach integrates predictions from all possible weights vectors over the posterior parameter distribution rather than use a single "optimal" set of network weights.
With integration a parameter vector that fits the data only slightly better than others will contribute only slightly more to prediction without exclusion of alternative models:

$$\hat{y}_{n+1} = \int y(x_{n+1}, w)\,P(w|(x_1, y_1), ..., (x_n, y_n))\,dw$$

Here with the notation(x_i, y_i) is represented the pair model input vs. observed values used to calibrate the model.
Integral involved in Bayesian prediction is generally hard – if not impossible - to evaluate analytically; numerical methods, namely Montecarlo methods, are available for this purpose.
The fundamental idea is to approximate $\int y(x,w)\,P(w|x,y)$ with the corresponding expected value $E[y(x,w)] \approx \frac{1}{n}\sum_{i=1}^{n} y(x,w_i)$ where w_i represents a sample of weight vectors generated from the distribution $P(w|x,y)$.

Metropolis Algorithm

This basic computation is inadequate for standard practical applications due to the multi – dimensionality and multi – modality in the posterior probability density for w; Metropolis algorithm [74] is an effective way to arrange for the distribution of weight vectors to correspond to $P(w|x,y)$.
Metropolis Algorithm generates candidate steps from Markov chains:

$$w_{new} = w_{old} + \varepsilon$$

but reject a proportion of the steps which lead to a reduction in the value of $P(w|x,y)$. This must be done with great care, however, in order to ensure that resulting sample of weight vectors represents the required distribution.
[1]

[1]The proof of convergence of Metropolis Hasting algorithm is far beyond the scope of this work; nevertheless it is possible to share an insight about the heuristic of the algorithm. The essential condition to generate samples from a target distribution is given by the "stationary balance" requirement; it means that, after a warm-up period, the random sequence reach the following state: N(x)p(x →y) = N(y) p(y →x) where:
N(x), N(y) = target distribution
p(x→y), p(y→x) = transition probability
The transition probability, due to the algorithm construction, is given by:
p(x →y) = a (x →y) acc (x →y)
p(y →x) = a (y→x) acc (y →x)

In the Metropolis algorithm this is achieved by using the following criterion:

If $P\left(w_{new}|\, x, y\right) > P\left(w_{old}|\, x, y\right)$ **Then** accept

Else accept with probability $\frac{P(\, w_{new}|x,y)}{P(\, w_{old}|x,y)}$

A critical point in Metropolis Algorithm and in any other MCMC samplers is the number of initial steps until the chain approaches stationarity; being that number related to the distance of starting value from the distribution's mode, the length of so named "burn-in period" can be reduced using gradient descent information to speed-up convergence.

---

where:
a (x →y), a (y →x) = probability of candidate y (x in the opposite case) coming from x (y in the opposite case)
acc (y →x), acc (x →y) = probability of "acceptance" of candidate x (y in the opposite case).
Choosing a (x → y) = a (y →x) the stationarity condition becomes:
N(x) acc (x→y) = N(y) acc (y→x)
that is:
acc (y→x) / acc (x→y) = N(x)/ N(y).
Now, because the "acceptance rule" of the algorithm:
If N(x)>N(y) then acc (y→x) = 1, acc (x → y) = N(y)/ N(x),
hence the stationarity condition is confirned; the same holds in the opposite case:
If N(x)<N(y) then acc (y→x) = N(x)/ N(y), acc (x → y) = 1

Chapter 7

Simulation results

The compulsory need of a model that describes a mean reverting process whose stationary distribution is designed to represent extreme values, comes from the failure of standard models in capturing the shocked dynamics of current recessional markets. Due to the relevance of public expectations in prices evolution, it is plausible that not only the financial crisis compels new mathematical tools but that obsolete models, driving misleading choices, compromise economic equilibrium.

Obviously, algorithms will not save the world but an honest enquiry about the conditions of their application could reveal useful insights about the relativity of classical economic "truths" first of all the assumption of invariance of economic rules or the dangerous idea of tendential infinite growth.

The forecasting of interest rates plays a fundamental role in global financial institutions investment strategies so as in central government macroeconomic interventions; to have a dimension of the importance of interest rates modeling, consider that specialized international companies earn millions of dollars selling "only" tables of numbers (interest rates curves) to professional investors.

A professional investor could be no so interested in the fact that predictions are good but that are the same on which its competitor takes crucial decisions: ideally, if all the market makers take the same decision, "predictions" become "prescriptions". So, "standard" models work well *because* they are *standard* until strong "structural" black outs (for example a pathologic money concentration that mismatches demand and offer) radically change the market scenarios.

A first (and reiterate) attempt to interpret market stresses - inspired by the silent "axiom" of economic rules invariance – is to force classical model parameters to include strong market anomalies; unfortunately, by this way, the model is tuned to expect exceptions as if they were normal facts so that the identifica-

tion of critical market periods is arbitrarily performed from the analyst during the calibration step rather than reached by a good model.

To test the usefulness of the model I'm proposing, I'll measure its success in elaborating VaR limits. I'll adopt an hypothesis test currently used in VaR evaluation to verify the effectiveness of the VaR limits obtained with the extreme value model and I'll compare the outcomes with those obtained from a traditional Vasicek approach.

7.1 Data

A real-world data panel is used for models testing, namely the Euribor *Eur001M* short rates assumed as a proxy of instantaneous rates; the choice of a five-years overall period (01/01/2007-31/12/2011) comes from a compromise between the need of a range wide enough to show radical changes in financial scenarios (in particular the 2008 drawdown) and the attempt to stay focused on an homogeneous frame, whose driving forces could be comparable under a macroeconomic point of view.
In order to evaluate the annual value at risk, the calibration dataset is segmented in windows of 252 observations each, and the outcomes of resulting model are "back tested" with 252 following observations not involved in calibration step.

	Descriptive statistics of data sample					
	2007	**2008**	**2009**	**2010**	**2011**	**2012**
n. of observations	255	256	257	257	257	227
mean	0,0396	0,0405	0,0074	0,0048	0,0102	0,002
std. deviation	0,0022	0,0054	0,0048	0,0015	0,0021	0,0014

7.2 Size of historical series

A central point in calibration task is the determination of the "right" portion of historical data that can be used to ensure a "good" prediction; so, we need to clarify univocally what the meaning of "good" in this context and how a data set could be theoretically "right" in order to support a prediction quantitatively marked as "good".
To assess the quality of a prediction the most intuitive thing to do is the measurement of the distance between the predicted values and the observed ones, as assumed by R^2, a common index belonging to the family of "goodness of fit"

Figure 7.1: *2007-2012 Eur001M curve - Source: Bloomberg L.P. Professional Services*

measures:

$$R^2 = 1 - \sum_{i=1}^{n} \frac{(y_i - f(x_i))^2}{(y_i - \bar{y})^2}$$

Here y_i is the observed value, $\bar{y}$ is the average of observed values and $f(x_i)$ is the predicted value; the prediction error $(y_i - f(x_i))^2$ is hence "normalized" on sample's standard deviation.

This seems a correct practice to evaluate the model error when the values we attempted to forecast are observable and comparable with predictions; but what do we know about the quality of the model "before" the "new history" happens ? What is the model's parameters set up that minimizes R^2 ?

A partial answer to these questions is embedded in the maximum likelihood criterion: if we maximize the probability of observed data *given* parameters, as in ML, we are reasonably confident to find representative parameters or, better, parameters are representatives if observed data are indeed the "most likely". But if no information about the parameter distribution is available, there is no indication about model generality and about its accuracy out of calibration

sample; to calibrate models taking into account parameter uncertainly we have to find the probability of parameters *given* the data through the Bayes rule:

$$P\left(\hat{\theta}\right) = P\left(\theta|\, y\right) \approx P\left(y|\theta\right) P\left(\theta\right)$$

The probability of a value $P\left(y\right)$ is, in coherence with the premises, given by the marginalization of many different likelihoods weighted for the respective probabilities:

$$P\left(y\right) = \int P\left(\hat{\theta}\right) P\left(y|\theta\right) d\theta$$

Suppose to define:

$$g\left(\theta\right) = \log\left(P\left(\hat{\theta}\right) P\left(y|\theta\right)\right) \Leftrightarrow P\left(\hat{\theta}\right) P\left(y|\theta\right) = \exp g\left(\theta\right)$$

expand $g\left(\theta\right)$ in Taylor series about the maximum $g\left(\theta_0\right)$:

$$\begin{array}{l} g\left(\theta\right) \approx g\left(\theta_0\right) + g'\left(\theta_0\right)\left(\theta - \theta_0\right) + \frac{1}{2} g''\left(\theta_0\right)\left(\theta - \theta_0\right)^2 = \\ = g\left(\theta_0\right) + \frac{1}{2} g''\left(\theta_0\right)\left(\theta - \theta_0\right)^2 \end{array}$$

(the first derivative is null due to the fact that it is evaluated at the maximum of the function)
Turning back to the definition of $P\left(y\right)$:

$$\begin{array}{l} P\left(y\right) = \int \exp\left(g\left(\theta_0\right) + \frac{1}{2} g''\left(\theta_0\right)\left(\theta - \theta_0\right)^2\right) d\theta \\ = \exp\left(g\left(\theta_0\right)\right) \int \exp\left(\frac{1}{2} g''\left(\theta_0\right)\left(\theta - \theta_0\right)^2\right) d\theta \end{array}$$

trasforming:

$$g''\left(\theta_0\right) = -\frac{1}{\sigma^2} \Leftrightarrow \sigma = -\sqrt{1/g''\left(\theta_0\right)}$$

from the definition of Gaussian distribution it is immediate that:

$$P\left(y\right) = e^{g(\theta_0)} \int e^{\frac{1}{2} g''(\theta_0)(\theta-\theta_0)^2} = e^{g(\theta_0)} \int e^{-\frac{(\theta-\theta_0)^2}{2\sigma^2}} = N\sigma\sqrt{2\pi}$$

$$= -\sqrt{1/g''\left(\theta_0\right)}\sqrt{2\pi} = \sqrt{\frac{2\pi}{-g''(\theta_0)}}$$

hence:

$$P\left(y\right) = \exp\left(g\left(\theta_0\right)\right) \sqrt{\frac{2\pi}{-g''\left(\theta_0\right)}}$$

This result offers a closed form approximation of the integral involved in $P\left(y\right)$ calculation; the method is due to Laplace.
Taking the logarithms and remembering previous definitions:

$$\log\left(P\left(y\right)\right) = \log P\left(\theta_0\right) + \log P\left(y|\theta_0\right) + \frac{1}{2}\log\left(\sqrt{2\pi}\right) - \frac{1}{2}\log\left(g''\left(\theta_0\right)\right) \qquad (1)$$

The last term in this equation reveals further details: the likelihood function in $g(\theta) = \log(P(\theta) P(y|\theta))$, can be written as $P(y|\theta) = \prod_{i=1}^{n} P(y_i|\theta)$, being the product of likelihood functions evaluated on ni.i.d. observations; it means that:

$$g''(\theta_0) = \frac{\partial^2 \log\left(P(\theta_0) \prod_{i=1}^{n} P(y_i|\theta_0)\right)}{\partial \theta_0} \propto \frac{\sum_{i=1}^{n} \partial^2 \log P(y_i|\theta_0)}{\partial \theta_0} \propto$$
$$\propto E\left\{\frac{\partial^2 \log P(y_i|\theta_0)}{\partial \theta_0}\right\} n$$

and that:

$$\frac{1}{2} \log(g''(\theta_0)) = \frac{1}{2} \log k + \frac{1}{2} \log n$$

Substituting in (1) and discarding $O(1)$ factors it's clear that:

$$\log(P(y)) = \log P(y|\theta_0) - \frac{1}{2} \log n$$

This approximation is called the Bayesian information criterion (BIC), and was first derived by Schwarz [81]; it can be used as a measure of the bias-variance tradeoff in model calibration and, for our purposes, offers a tool to estimate$n$, the recommended size of historical series intimately related to the sampling error.

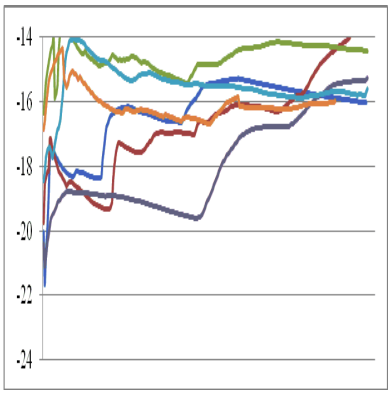

Figure 7.2: *BIC measures for each annuity*

Observing the BIC values related to the sample sizes for each of the year in scope, it becomes clear that along an under-sampling window the indicator decreases till a minimum that is the starting point of an over-sampling frames; it means that the information gain progressively induced by the sample growth, compensates only partially the increase in overall volatility.
The estimation period is then fixed from the starting time to the last time step before the BIC value restart growing; in all the data samples the estimation period is quite similar – around 100 observations – so that I preferred, with an acceptable simplification, to work with a size of 100 observations for each sampled year.

7.3 Parameters optimization

Parameters calibration is achieved through least square method implemented with the Levenberg-Marquardt optimization algorithm (see chapter 6 for a review); convergence outcomes related to the in-sample error are plotted.

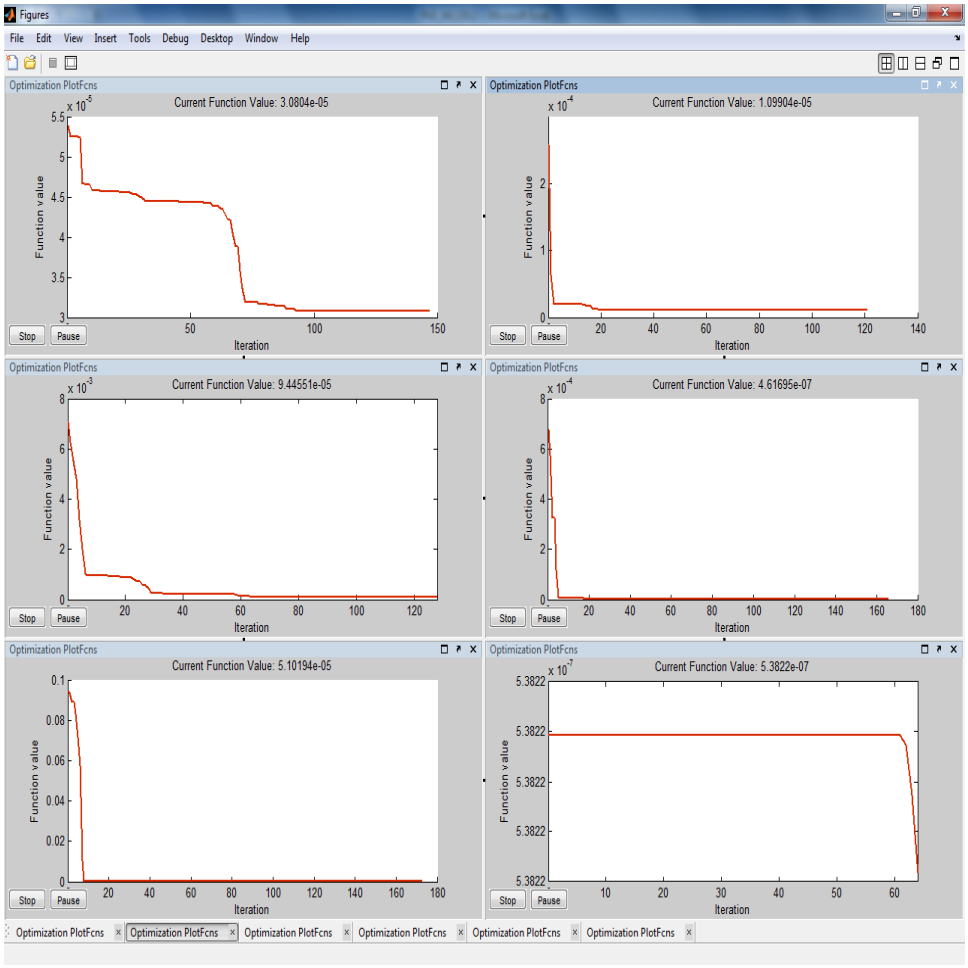

Figure 7.3: *In-sample fitting error*

The values of estimated parameters are reported in conjunction with respective confidence intervals; to compute standard error the distribution of each of the model parameters is simulated numerically as follows.
After the parameters have been estimated – call these "master" parameters -, simulate a sample path of length equal to the historical sample from the model configured with those parameters. For this sample path one then estimates the

best fit parameters as if the simulated path were the historical one, and obtains a new sample of the parameters. Then again, iterating simulation from the "master" parameters, one goes through this procedure and obtains other samples of parameters; doing so, one builds a random distribution of each parameter, all starting from the master parameters.

Table: EVT-Vasicek parameters statistics

		k	c	h
2007	*mean*	1,25727E-09	3,59202E-12	4,4E-07
	st.dev	2,27018E-09	5,38853E-12	3,84E-07
2008	*mean*	3,04897E-08	1,19074E-09	3,3E-06
	st.dev	6,9729E-08	2,92584E-09	2,99E-06
2009	*mean*	5,39532E-06	5,89557E-07	0,000172
	st.dev	2,11523E-05	1,76916E-06	0,000176
2010	*mean*	1,34248E-12	4,55811E-15	6,48E-09
	st.dev	5,02212E-12	1,65444E-14	5,74E-09
2011	*mean*	1,5289E-09	8,03523E-12	3,54E-07
	st.dev	2,08204E-09	1,09603E-11	2,97E-07
2012	*mean*	5,49374E-08	6,793E-10	4,09E-06
	st.dev	8,63231E-08	1,25389E-09	3,25E-06

7.4 Simulation outputs

After model's parameters estimation model's outcomes are finally obtained. The set of following figures depicts, for each annuity:

1. Montecarlo rates paths diffusion for classical Vasicek model
2. Montecarlo rates paths diffusion for revised EVT- Vasicek model
3. In sample & out of sample fitting of the calibrated model
4. Empirical distribution of rates at the end of the simulation period for both models.

The empirical distribution of rates is represented here to compare Vasicek and revised EVT-Vasicek models in terms of extreme loss forecasting, what risk management pratictioners fundamentally need to prevent portfolio crashes; this kind of model assessment is analytically developed through the Kupiec Test.

7.4.1 Kupiec Test

VaR backtesting requires a method to verify the reliability of model outputs when the model is applied on samples selected outside the calibration domain.

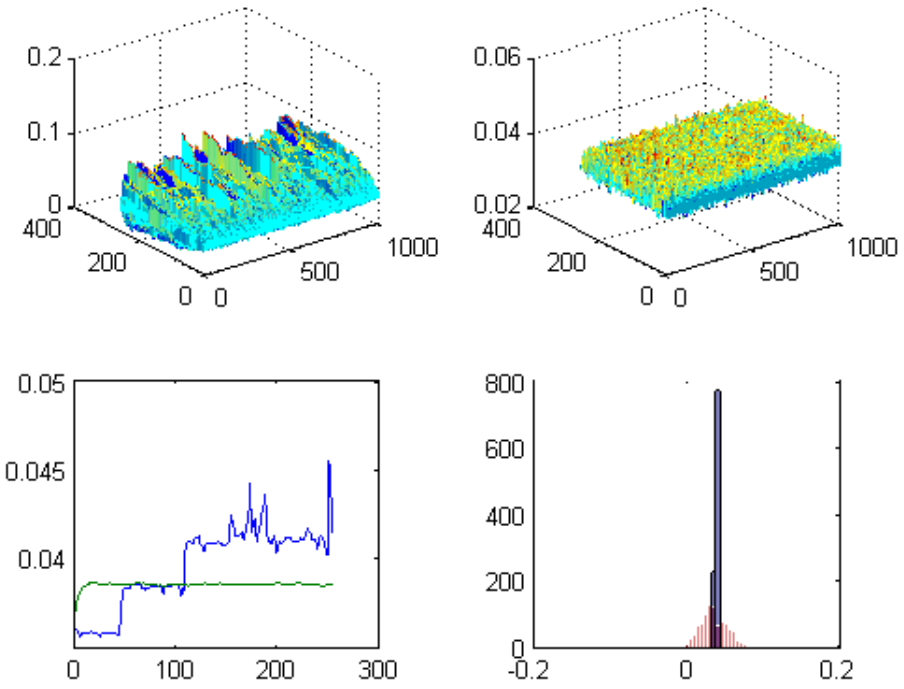

Figure 7.4: *2007 simulation output*

Being VaR the worst loss in a time horizon for a given confidence limit, the Kupiec test [57] is an hypothesis test that assesses the compatibility of the number of realized VaR violations with that limit.
More formally:
Let be x the effective number of threshold crossing (failures) on n values and c the confidence level; the null hypothesis is that frequency of failures is equal to the predicted failure rate p of the model.
Generally the probability of x violations out of n total trials is a function of the possible combinations of violations:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{x-n}$$

The log-likelihood ratio statistic is an index of the gap between the theoretical amount and the observed one of the VaR violations:

$$LR = -2\ln\left[(1-p)^{x-n} p^n\right] - 2\ln\left[(1-{}^{x}\!/_{n})^{x-n} ({}^{x}\!/_{n})^n\right]$$

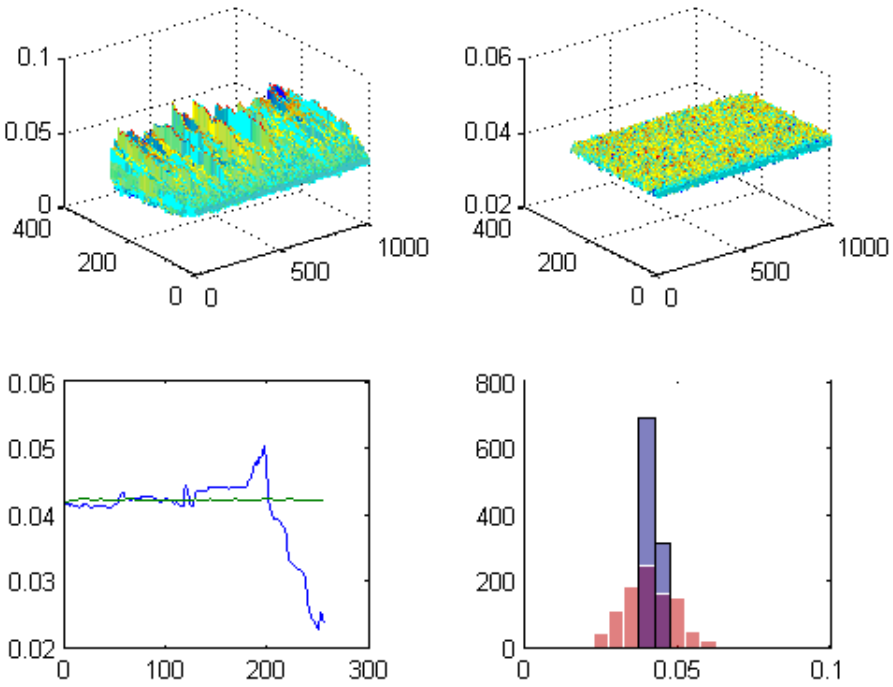

Figure 7.5: *2008 simulation output*

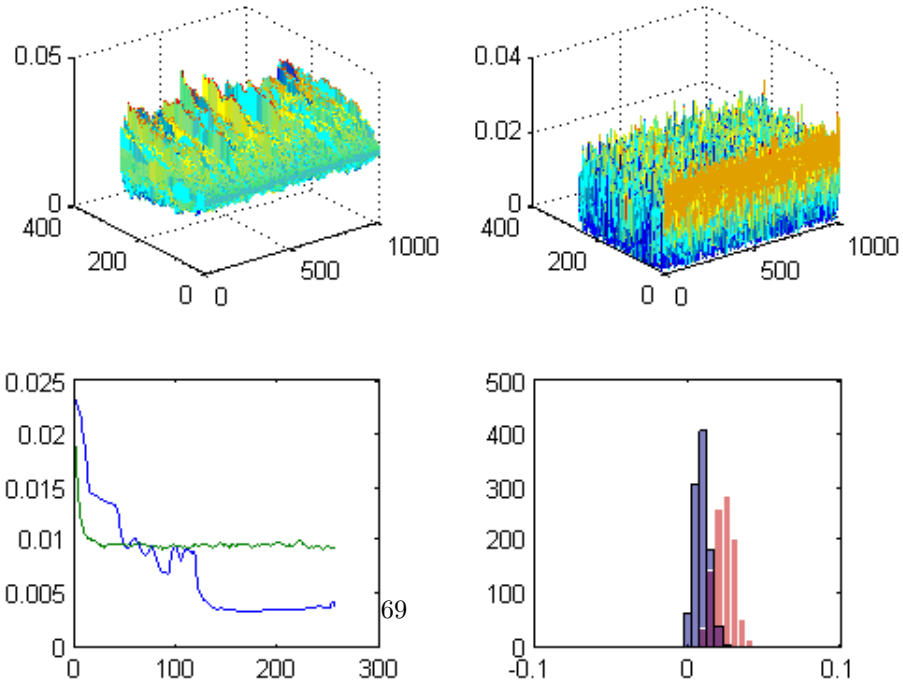

Figure 7.6: *2009 simulation output*

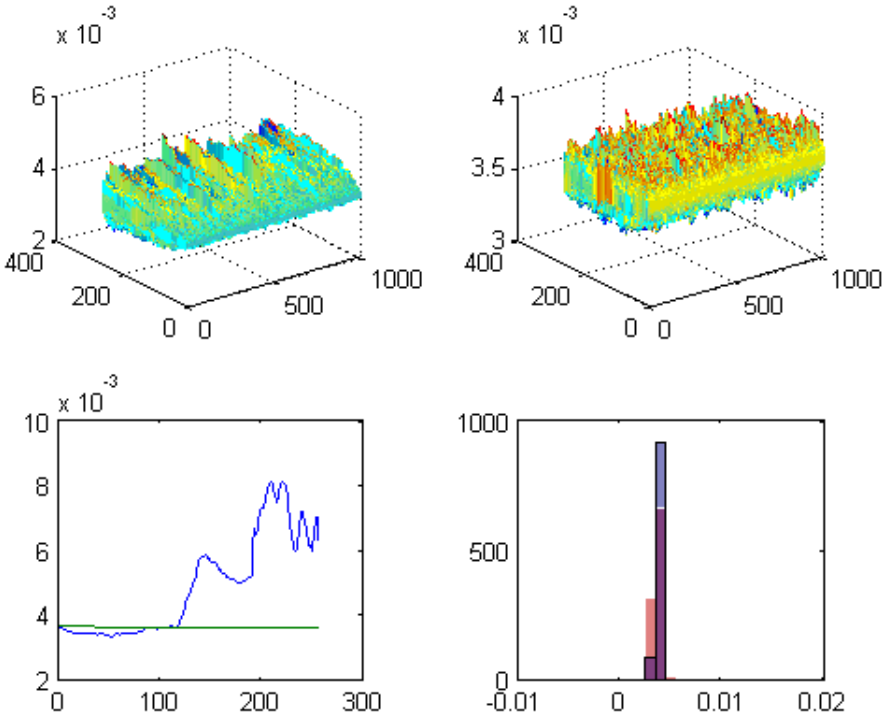

Figure 7.7: *2010 simulation output*

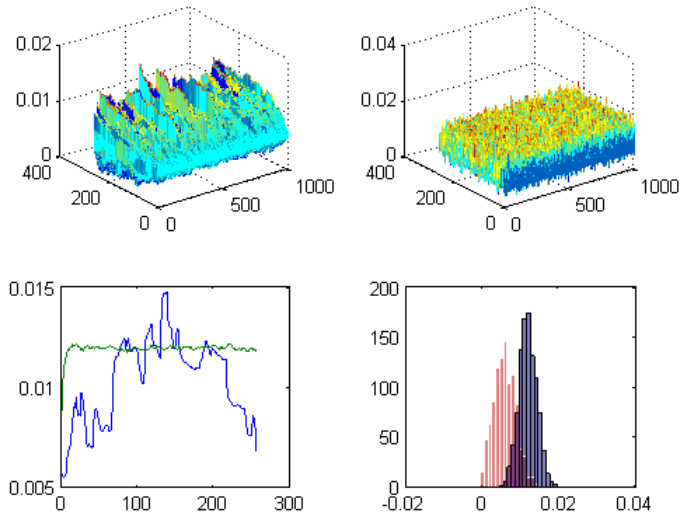

Figure 7.8: *2011 simulation output*

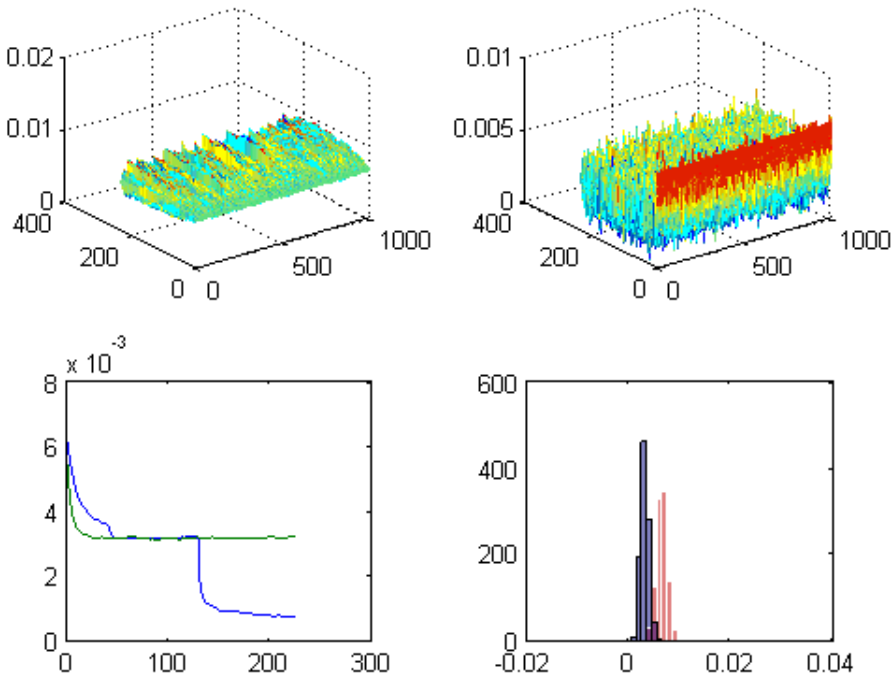

Figure 7.9: *2012 simulation output*

Its distribution, for $n \to \infty$ and under the null hypothesis$p = \frac{x}{n}$, is a chi-squared with one degree of freedom.
The acceptance regions for the confidence level c are derived from the chi-squared percentiles: $\{x : LR \leq c\}$
If the statistic LR overlays the acceptance limit the null hypothesis $p = \frac{x}{n}$ is rejected: the VaR cannot be confirmed. In what follows the Kupiec test outputs are displayed.

Table: Kupiec test (p-value = 0,1 - critical value = 2,705)

		2007	*2008*	*2009*	*2010*	*2011*	*2012*
Failures	*Vasicek*	22	31	7	136	0	6
Failures	*EVT Vasicek*	0	0	0	127	9	0
LR	*Vasicek*	37,179	61,679	5,4380	473,36	1,1102	4,0060
LR	*EVT Vasicek*	1,1102	1,1102	1,1102	431,49	8,670	1,1102
p-value	*Vasicek*	1,08E-09	4,04E-15	0,0197	5,9E-105	0,2920	0,0453
p-value	*EVT Vasicek*	0,2920	0,2920	0,2920	7,68E-96	0,0032	0,2920

As we could expect the Vasicek model is severely penalized by the market turbulence and only in one time window leads to null hypothesis acceptation.Revised Vasicek works better in all the sessions except the 2011 forecast where 9 violations occur and the likelihood ratio passes the critical value causing the hypothesis rejection; the only further macroscopic failure of the revised model is in the year 2010, whose market dynamics are evidently hard to capture.The anomalous behavior of these dynamics implies now expensive financial troubles but can be an interesting source of new information about the microstructure of hazardous economic scenarios.

Chapter 8

Conclusions

The market crash started (revealed) in 2008 is the definitive check point for standard financial quantitative models; they indubitably fall in capturing adverse scenarios and neglect extreme events because they are based on distributional hypothesis that are systematically violated in the real life.
On the other hand, if a model, or a set of models, is inadequate to represent market movements, there is a low chance to make profitable business avoiding fatal losses: the risk underestimation increases the default probability.
From a practical point of view it is hence necessary to adopt new financial models that relax structural constraints related to the distribution of observed prices; this requirement leads to the theoretical enquiry about the relationship between the characteristic of a stochastic process and the distribution of the random variable subject to that process.
My research applies standard results of extreme value theory to stochastic processes in order to modify the distributional assumptions of a classical interest rate model and to adapt it to the representation of extreme returns; while this is the main task of this work, it can inspire most general considerations and contributions as I will suggest later, after a comment on the experimental results obtained in the previous chapter.

8.1 Notes about the outcomes

The main risk embedded in my research is due to the disproportion between the importance of the topic – the market data behavior in stressed economic scenarios – and the resources – necessarily limited – I have at my disposal to analyze the same topic; this risk consists of a misevaluation of the problem because the partiality of the solution I propose, while it should be useful to empower the effort on the analysis to compensate the imperfection of *my* analysis.

At a first glance, looking at the test on VaR threshold violation, it seems confirmed the appropriateness of an approach that takes into consideration the

fatness of distribution tails and pays a particular attention to the volatility dynamics: this real market challenge of the EVT-Vasicek model is quite encouraging, and the model is currently adopted for experimental purposes from the model validation unit of an Italian top insurance group.

We could criticize that the model works well during economic recessions but the alert about an upcoming crisis should be the output of the model, not a working condition; in other words the model must work well in every set of circumstances, being able to modify its output so to anticipate real dangers. The adjective "real" is not pleonastic here, it recalls the fact that warnings about "false" dangers have to be avoided, and that the augmentation of false positive to reduce the false negative is not a best practice; so, what I planned to achieve is a satisfactory "reactivity" of the model, its attitude to increment VaR only when market is deteriorating (a model that increases VaR in an indiscriminate way is simply the classical model with inflated variance).
In EVT-Vasicek approach, the volatility of the process is a function of the interest rate; this characteristic makes the same process sensitive to the updated market level, as I intended.

The experimental session I ran to reach these results persuaded me about the "price" we have to pay to increase the reactivity in the sense just mentioned: a factual renounce to model adherence to distributional central values.
The particular structure of the sigma function I derived imposes a tradeoff between a good drift fitting and a good volatility fitting: if parameters are optimized to follow the central tendency, the model presents a too strong heteroskedasticity; conversely, when calibrated with respect to volatility – and adapted to theoretic drift constraints – the model minimizes the drift contribution. The analysis of this tradeoff could be developed in terms of bias-variance balance and included in an upcoming research plan related to this first exploration.

8.2 Further developments

The most obvious direction of next research steps, implicit in the current research activity, concerns a more comprehensive set of tests regarding interest rates models – not only the Vasicek approach – but, due to the strong analogies with the aspects treated in the previous chapters, specifications can be omitted here as superfluous; more interesting insights come, instead, from an higher level of generality of the approach I'm proposing.

Despite the fact that the model I experimented here is specialized on interest rates given a specific stationary distribution, the result described in chapter 5 can be referred as "general": it is derived independently from the detail of the distribution and of the process used to build a model adapted to a real market situation; that result, offers a tool to write down the SDE of a process

given it's stationary distribution. This tool could hence represent a resource to inspect the perspective from which it could be possible to overcome the limits imposed to the quantitative finance by unrealistic distributional assumptions.

The two mainstreams of possible improvements of this research are related to the two "degree of freedom" admitted in specializing the general result: we can work on the distributional foundations to find and justify appropriate probability laws, or, on the other side, we can work on many stochastic processes related to many financial dynamics and, potentially, to every standard financial model.

8.2.1 Loss distribution and market microstructure

The choice of an adequate probability law to represent a random variable distribution should derive from an arguable theory about market information arrival and price formation.
As introduced in the first chapter, there are paradigmatic examples in Fama, Mandelbrot, Clark, about the analysis of market behavior in order to explain the price (or return, or losses) distribution.
Let me quote some sentences in Fama [29]:
"The random walk theory is based on two assumptions: 1) price changes are independent random variables, and 2) the changes conform to some probability distribution...
This paper will be concerned with the nature of the distribution of price changes.
... the usual assumption, which we shall henceforth call the Gaussian hypothesis, was that the distribution of price changes in a speculative series is approximately Gaussian or normal...
If the price changes from transaction to transaction are independent, identically distributed, random variables with finite variance, and if transactions are fairly uniformly spaced through time, the central-limit theorem leads us to believe that price changes across differencing intervals such as a day, a week, or a month will be normally distributed since they are simple sums of the changes from transaction to transaction.
... it has been found that the extreme tails of empirical distributions are higher (i.e., contain more of the total probability) than those of the normal distribution.
... the stable Paretian hypothesis, makes two basic assertions: 1) the variances of the empirical distributions behave as if they were infinite, and 2) the empirical distributions conform best to the non-Gaussian members of a family of limiting distributions which Mandelbrot has called stable Paretian...
The property of stability or invariance under addition is responsible for much of the appeal of stable Paretian distributions as descriptions of empirical distributions of price changes"

This citation sketches some elements that are useful to setup the analysis of market microstructure: in brief, Mandelbrot needs to take into account big

deviations from normality while preserving the idea of price changes as "summation of price changes from transaction to transaction" and hence the property of "stability or invariance under addition" for the price distribution.
The Paretian distribution both represents fat-tails and satisfies the stability need, but its infinite variance has been considered a shortcoming from subsequent authors that proposed alternative hypothesis as the mixture distribution hypothesis met in the first chapter.
In those pages, I highlighted the max stability property of the Extreme Values distribution family, analogous to the alpha-stability property of Paretian distribution; in particular it can be proved that if a random variable is in the maximum domain of attraction of a Frechet type extreme value distribution, then it is in the domain of attraction of a Gaussian distribution that, for the central limit theorem, is stable under addiction.
This fact makes the Frechet distribution a good candidate as limit distribution for market information; it fulfills both the basic requirements previously indicated as essential for such a distribution: the closure with respect to the sum and fat tails without the drawback of infinite variance.

8.2.2 Standard models revised

The analysis of market microstructure can guide us in selecting the most appropriate probability law for market information; given an appropriate law we have the opportunity to describe processes related to specific models we need.
A second research area interested by the change of distributional assumption of financial stochastic models is hence more applicative and – in its widest sense – involves all the traditional milestones of quantitative finance.
The Black & Scholes option pricing model – Nobel prize in 1997 - is the archetype for many of successive contribution to quantitative finance; its reinterpretation under changed distributional hypothesis represents an interesting applicative example and opens a valuable scenario with respect to current business priorities.
As I observed about the Vasicek model extensions (cfr. chapter 5), even for Black & Scholes updating I didn't find specific researches treating the explicit link between stochastic processes and extreme value theory; nevertheless there are some works that can be reputed affine to the topic discussed here and can support its comprehension.
An instructive paper from Moriconi [76] proposes an abstraction of the Black & Scholes pricing formula derived independently from distributional constraints; this result is influenced by a previous contribution from L.Borland [12] in which price fluctuations in the option pricing model follow a nonlinear Fokker-Planck equation which maximizes the Tsallis nonextensive entropy.
A martingale-equivalent approach to an extreme-values variation of the Black & Scholes model is developed by Markose and Alerton [68]; their paper preliminarily illustrates a wide repertory of different methods used to vary distributions

implied in option pricing models that have arisen since the early works, but, curiously, falls to remember the famous intervention of Merton [73].
Merton, whose name is strictly associated with Black & Scholes for merits in option pricing studies and for the default estimation models we'll see further, has, himself, released the most quoted application of jump-diffusion processes to option pricing formula.
The Black & Scholes formula is fundamental not only for its pricing purpose but, at least, for two more reasons: the first is historical, it made the Ito's calculus the standard for the quantitative finance; the second is connected with a current and urgent problem to solve, the problem to know effectively a market counterparty and its actual solvency profile.
In 1974 R.C. Merton [72] proposed an enhancement of the Black & Scholes formula arranged to estimate the default probability of a public company through the price of its equity; at present, such a tool – adapted to the existing distributional constraints – offers a valid alternative to the official ratings supplied by agencies, whose judgment can be affected by clash of interest.
To complete the exemplification of the possible extensions of standard models, the Merton model will be sketched here in its extreme value version, i.e., assuming a GEV pdf as stationary distribution of the option price stochastic process.
In Merton's model an equity of a public company is considered as an option written on the company's assets. Therefore if the asset value in t is lower than the liabilities amount at the same timestep, the company goes bankrupt and must repay the debts; the investor, instead, loses, at worst, the entire invested capital.
The equity E in time T has value:

$$E_T = \max\left(V_T - D, 0\right)$$

where:
V_T is the assets value at T
D is the liabilities value at T

Applying the EVT modified Black & Scholes formula (see [68]), it is:

$$E_T = e^{-r(T-t)}\left\{V_t\left(\left(1-\mu_V+\sigma_V/\xi\right)e^{-H^{-1/\xi}} - \frac{\sigma_V}{\xi}\Gamma\left(1-\xi,-H^{-1/\xi}\right)\right) - Ke^{-H^{-1/\xi}}\right\} \tag{8.1}$$

with:

$$H = 1 + \frac{\xi}{\sigma_V}\left(1 - \frac{D}{V_t} - \mu_V\right)$$

$\Gamma\left(\alpha, x\right)$ is the incomplete Gamma function
μ_V= the assets mean
σ_V= the assets standard deviation
ξ = GEV shape parameter
$e^{-H^{-1/\xi}} = P\left(K \middle| K < V\right)$ is the company's default probability.

Be aware of $E_i(\xi)$ is function of V_i but it isn't directly observable; for this reason V_i for $i > 0$ is recursively obtained as follows: $V_i = V_{i-1} + \frac{E_i}{\partial E/\partial V}$ being

$$\frac{\partial E}{\partial V} = e^{-Q^{-\omega}} Q^{-1-\omega} \frac{K}{V} \left(\frac{V(1-\mu+\omega\sigma) - K}{V\sigma} - \omega Q \right) - \sigma\omega\Gamma\left(1 - \frac{1}{\omega}, Q^{-\omega}\right) + e^{-Q^{-\omega}} (1-\mu+\omega\sigma),$$

$$\omega = \frac{1}{\xi}, Q = 1 + \frac{(1 - K/V - \mu)}{\omega\sigma}$$

The value of V_0 must be initialized recurring to a further informative condition derived by the assets process $dV = \mu_V V dt + \sigma_V V dW$ through the Ito's lemma:

$$dE = \left(\frac{1}{2}\sigma_V^2 V^2 \frac{\partial^2 E}{\partial V^2} + \mu_V V \frac{\partial E}{\partial V} + \frac{\partial E}{\partial t} \right) dt + \sigma_V V \frac{\partial E}{\partial V} dW$$

V_0 is hence deducted from the diffusion term $\sigma_E E_0 = \frac{\partial E}{\partial V}\sigma_V V_0$

Bibliography

[1] Bachelier, L. - Theorie de la speculation, Annales de l'Ecole Normale Superieure 17, 21-86. - 1900

[2] Bali, T. G. - An Extreme Value Approach to EstimatingVolatility and Value at Risk. Journal of Business, 76, 83-108 - 2003

[3] Bali, T. G., Neftci, S. - Disturbing Extremal Behavior of Spot Rate Dynamics. Journal of Empirical Finance, 10, 455-77 - 2003

[4] Balkema, A.A., de Haan, L.,. Residual lifetime at great age. Annals of Probability 2, 792-804. - 1974

[5] Barndorff-Nielsen, O.E. - Normal inverse Gaussian distributions and the modelling of stock returns, Scandinavian Journal of Statistics 24, 1-13 - 1997

[6] Barndorff-Nielsen, O. E. - Processes of normal inverse Gaussian type. Finance Stoch. 2, 41-68 - 1998

[7] Barndorff-Nielsen, O. E. and Shephard, N. - Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics (with discussion). J. Roy. Statist. Soc. Ser. B 63 , 167-241 - 2001

[8] Bibby, B. M., Sorensen M. - A Hyperbolic Diffusion Model for Stock Prices, Finance Stoch. 1, 25-41 - 1997

[9] Black, F., and Karasinski, P. - Bond and option pricing when short rates are log-normal, Financial Analysts Journal July-Aug, 52-29 - (1991)

[10] Black,F. and Scholes,M. - The Pricing of Options and Corporate Liabilities - Journal of Political Economy, Vol. 81, pp. 637-659, 1973

[11] Borkovec, M. Kluppelberg,C. - Extremal behaviour of diffusion models in finance, Stochastic Processes and their Applications, 117, 432-456 - 2007

[12] Borland, L.- A Theory of Non-Gaussian Option Pricing - 2008

[13] Clark, P.K. - A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, Econometrica, 41, 135-155 - 1973

[14] Coles, S. G. - An Introduction to Statistical Modeling of Extreme Values. London: Springer - 2001

[15] Cont, R., Tankov, P. - Financial Modelling with Jump Processes. Chapman & Hall - 2004.

[16] Cont, R. - Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance Volume 1, 223-236, Institute Of Physics Publishing - 2001

[17] Cootner, P. ed. -. The Random Character of Stock Prices. Cambridge, Mass.: M.I.T. Press. - 1964
[18] Kou, S. G.: A jump diffusion model for option pricing. Manag. Sci. 48, 1086-1101 - 2002
[19] Cox, J. Ingersoll, J.E., Ross. S.A. - A theory of the term structure of interest rates. Econometrica, 53:385-402 - 1985
[20] Dacorogna, M. M., Pictet, O. V. Muller, U. A., de Vries, C. G. - Extremal forex returns in extremely large data sets, Extremes - 2001
[21] Davis, R. - Maximum and minimum of onedimensional diffusions, Stochastic Processes and their Applications, 13, 1-9 - 1982
[22] De Haan, L. Ferreira,A.- Extreme value theory. An introduction - Springer, 2006
[23] Dornbusch, R. - A primer on emerging market crisis, manuscript, MIT, Department of Economics - 2001
[24] Dupire, B. - Pricing and Hedging with Smile - Risk Magazine, 1993
[25] Embrechts, P. - Extremes and Integrated Risk Management, Risk Books and UBS Warburg, London. - 2000
[26] Embrechts, P. - Extreme value theory in finance and insurance, manuscript. Department of Mathematics, ETH, Swiss Federal Technical University. - 1999
[27] Embrechts P., Kluppelberg,C. Mikosch,T.- Modelling Extremal Events for Insurance and Finance - Springer-Verlag - 1997
[28] Embrechts P., S. Resnick and G. Samorodnitsky.- Extreme value theory as a risk management tool North American Actuarial Journal 3, 30-41-1999
[29] Fama, E. F.- Mandelbrot and the Stable Paretian Hypothesis - The Journal of Business, Vol. 36, No. 4 pp. 420-429, University of Chicago Press, 1963
[30] Feller, W. - An Introduction to Probability Theory and its Applications, Vol. II. New York: Wiley - 1971
[31] Fernandez , V. - Risk management under extreme events. International Review of Financial Analysis 14 113-148 - 2005
[32] Fisher, R.A. and Tippett, L.H.C. - Limiting forms of the frequency distributions of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society, 24, 180-190 - 1928.
[33] Fokker. A.D. - Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld - Ann. Phys. 348 (4. Folge 43), 810 - 820 - 1914
[34] Frechet, M . - Sur la loi de probabilite de l'ecart maximum, Ann. Soc. Polon. Math. (Cracovie), 6, 93-116 - 1927
[35] Gencay, R. Selcuk, F. Ulugulyagci - High volatility, thick tails and extreme value theory, Insurance: Mathematics and Economics.33 337-356 - 2003
[36] Gencay, R. Selcuk F.- Overnight borrowing, interest rates and extreme value theory - European Economic Review, 50 547-563 - 2006
[37] Glasserman, P.- Monte Carlo Methods in Financial Engineering - Springer, 2004
[38] Gnedenko, B. V. - Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math., 44, 423-453. - 1943
[39] Gumbel, E.- Statistics of Extremes. Colombia University Press, New York. - 1958

[40] Harrison, J. M. Pliska S. R - Martingales and stochastic integrals in the theory of continuous trading. Stochastic Processes and their Applications 11, 215-260 - 1981
[41] Hosking, J.R.M., Wallis, J.R.- Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29, 339-349.- 1987
[42] Heath,D., Jarrow,R., Morton, A. Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation Econometrica, 60, pp. 77 - 105 - 1992
[43] Hill, B. M. - A simple general approach to inference about the tail of a distribution, Annals of Statistics, 3, 1163-1174. - 1975
[44] Ho, T. S. Y. , Lee S.B. - Term structure movements and pricing interest rate contingent claims. Journal of Finance 41, 1011-1029 - 1986
[45] Hols, M. C., de Vries, C. G. - The limiting distribution of extremal exchange rate returns, Journal of Applied Econometrics, 6, 287-302. - 1991
[46] Hull, J. C. and A. White - Pricing interest rate derivative securities. Review of Financial Studies 3, 573_592. - 1990
[47] Hull, J.C. - Options, futures and other derivatives, 6th edition. Prentice Hall - 2006.
[48] Jackwerth,J.C. Rubinstein,M.- Recovering probability distributions from option pricing - The Journal of Finance, Vol. 51, N. 5, pp. 1611-1631, 1996
[49] Ito, K. - Multiple Wiener integral. J. Math. Soc. Japan 3, 157-169 - 1951
[50] Jondeau,E. Poon,S. Rockinger, H.M.- Financial Modeling Under Non-Gaussian Distributions - Springer 2007
[51] Kac, M. - On Some Connections between Probability Theory and Differential and Integral Equations, Proc. Second Berkeley Symp. on Math. Statist. and Prob. - 1951
[52] Karatzas, I. and. Shreve, S.E - Brownian Motion and Stochastic Calculus. Springer - 1991
[53] Karlin, S., Taylor, M. - A Second Course in Stochastic Processes. Orlando, FL: Academic Press - 1981
[54] Kloeden, P.E. Platen,E.- Numerical solution of stochastic differential equations - Springer-Verlag 1992
[55] Kluppelberg, C. - Risk Management with Extreme Value Theory - Sonderforschungsbereich 386, Paper 270, 2002
[56] Kotz, S., Nadarajah, S. - Extreme Value Distributions: Theory and Applications. London: Imperial College Press. - 2000
[57] Kupiec, P. H. - Techniques for verifying the accuracy of risk management models - Journal of Derivatives, 3:73-84, 1995
[58] Kusche, T.- Modeling financial markets with extreme risk - Universitaat Regensburg Mathematik, 2008
[59] Leadbetter, M. R. Lindgren, G. and Rootzen. H. - Extremes and Related Properties of Random Sequences and Processes. Springer - 1983
[60] Levenberg, K. - A Method for the Solution of Certain Non-Linear Problems in Least Squares . The Quarterly of Applied Mathematics, 2: 164-168 - 1944
[61] Levy, P. - Calcul des Probabilits. Gauthier-Villars, Paris -1925

[62] Longin, F.M. - From Value at Risk to Stress Testing: The Extreme Value Approach. Journal of Banking and Finance, 24, 1097-130 - 2000
[63] MacKay, D. - Introduction to Monte Carlo methods. M. Jordan, editor, Learning in Graphical Models, pages 175-204. The MIT Press, Cambridge, MA, USA, - 1999
[64] Mc Kean, H.P. - Stochastic Integrals, Academic Press - 1969
[65] Madan, D. Roynette, B. Yor, M. - Option Prices as Probabilities. Finance Research Letters.5,79-87. - 2008
[66] Mandelbrot, B. - The variation of certain speculative prices - Journal of Business, 36:394-419, 1963
[67] Marquardt,D.W. - An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, 11:431-441 - 1963
[68] Markose,S. Alentorn,A.- The Generalized Extreme Value (GEV) Distribution, Implied Tail Index and Option Pricing - , 2005
[69] McNeil,J. - Extreme Value Theory for Risk Managers - Internal Modelling and CAD II published by RISK Books, 93-113 - 1999
[70] McNeil, A. J., Rudiger Frey. - Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. Journal of Empirical Finance, 7, 271-300 - 2000
[71] Merton, R. - The Theory of Rational Option Pricing , Bell Journal of Economics and Management Science. 4, pp. 141-183. - 1973
[72] Merton, R.C. - On the pricing of corporate debt: The risk structure of interest rates - The Journal of Finance, Vol. 29, No. 2, pp. 449-470 - 1974
[73] Merton, R. C.- Option pricing when underlying stock returns are discontinuous. Journal Financ. Econ. 3 125-144. - 1976
[74] Metropolis, N. Rosenbluth, A.W. Rosenbluth, M.N. Teller, A.H. and Teller,E. - Equations of State Calculations by Fast Computing Machines . Journal of Chemical Physics, 21:1087-1092, - 1953
[75] Mikosch,T.- Regular Variation, Subexponentiality and Their Applications in Probability Theory
[76] Moriconi,L. - Delta Hedged Option Valuation with Underlying Non-Gaussian Returns -
[77] Oksendal, B.- Stochastic Diferential Equations - Springer - Verlag, 2000
[78] Parkinson, M. - The Extreme Value Method for Estimating the Variance of the Rate of Return, The Journal of Business, Vol. 53, No. 1 (Jan., 1980), pp. 61-65 The University of Chicago Press - 1980
[79] Pickands, J.- Statistical inference using extreme order statistics. Annals of Statistics 3, 119-131.- 1975
[80] Planck, M., - Sitz. ber. Preuss. Akad. Wiss, 324 - 1917
[81] Resnick, S. I. - Extreme Values, Regular Variation, and Point Processes. Springer - 1987
[82] Schwartz, G.- Estimating the dimension of a model - The Annals of Statistics, Vol. 6, N. 2, 461-464, 1978
[83] Vasicek, O.- An Equilibrium Characterization Of The Term Structure - Journal of Financial Economics 5, 177-188 - 1977

[84] Von Mises, R. - La distribution de la plus grande de n valeurs. Reprinted in Selected Papers Volumen II, American Mathematical Society, Providence, R.I., 1954, pp. 271-294 - 1936